DEVELOPMENT OF A DOMAIN ANALYSIS MODEL FOR ELECTRONIC INSTITUTIONAL REVIEW BOARD SYSTEMS: A FEASIBILITY STUDY

by

Shan He

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biomedical Informatics

The University of Utah

August 2013

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Shan He**

has been approved by the following supervisory committee members:

John F. Hurdle	, Chair	**5/7/2013** Date Approved
Scott P. Narus	, Member	**5/7/2013** Date Approved
Julio Cesar Facelli	, Member	**5/7/2013** Date Approved
Lee Min Lau	, Member	**5/7/2013** Date Approved
Jeffrey R. Botkin	, Member	**5/7/2013** Date Approved

and by **Julio Cesar Facelli** , Chair of

the Department of **Biomedical Informatics**

and by Donna M. White, Interim Dean of The Graduate School.

ABSTRACT

Clinical research plays a vital role in producing knowledge valuable for understanding human disease and improving healthcare quality. Human subject protection is an obligation essential to the clinical research endeavor, much of which is governed by federal regulations and rules. Institutional Review Boards (IRBs) are responsible for overseeing human subject research to protect individuals from harm and to preserve their rights. Researchers are required to submit and maintain an IRB application, which is an important component in the clinical research process that can significantly affect the timeliness and ethical quality of the study. As clinical research has expanded in both volume and scope over recent years, IRBs are facing increasing challenges in providing efficient and effective oversight. The Clinical Research Informatics (CRI) domain has made significant efforts to support various aspects of clinical research through developing information systems and standards. However, information technology use by IRBs has not received much attention from the CRI community.

This dissertation project analyzed over 100 IRB application systems currently used at major academic institutions in the United States. The varieties of system types and lack of standardized application forms across institutions are discussed in detail. The need for building an IRB domain analysis model is identified.

.

In this dissertation, I developed an IRB domain analysis model with a special focus on promoting interoperability among CRI systems to streamline the clinical research workflow. The model was evaluated by a comparison with five real-world IRB application systems. Finally, a prototype implementation of the model was demonstrated by the integration of an electronic IRB system with a health data query system.

This dissertation project fills a gap in the research of information technology use for the IRB oversight domain. Adoption of the IRB domain analysis model has potential to enhance efficient and high-quality ethics oversight and to streamline the clinical research workflow.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Clinical research is aimed at investigating new approaches to the diagnosis, treatment, and prevention of human disease. It plays an important role in improving the quality of healthcare. Clinical researchers are faced with increasingly complex workflows through the entire study life cycle and critical requirements for efficient management and reuse of large amounts of research data and metadata. Significant effort has focused on designing biomedical informatics tools that support efficiently conducting and tracking studies. These efforts led to the emergence of a domain that has become a subdiscipline of biomedical informatics focused on clinical research referred to as Clinical Research Informatics (CRI).[1,2] Example CRI systems include but are not limited to Secondary Use of Health Data systems (SUHD) (e.g., hospital data warehouses, distributed health data repositories, disease registries), Clinical Trial Management Systems (CTMS), Electronic Data Capture (EDC) systems (e.g., REDCap,[3] OpenClinica[4]), Clinical Trial Registries (CTR) (e.g., ClinicalTrials.gov), and electronic submission and reporting systems (e.g., Electronic Submissions Gateway,[5] MedWatch[6]) supported by the Food and Drug Administration (FDA).

Human subject protection is an obligation essential to the clinical research endeavor, much of which is governed by regulations and rules such as the Department of Health

and Human Services' (HHS) Protection of Human Subjects regulations[7]; the Food and Drug Administration's (FDA) human subject protection regulations[8–11]; the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule (also known as Standards for Privacy of Individually Identifiable Health Information); and the most recent Omnibus Final Rule which is a modification to the HIPAA rules under the Health Information Technology for Economic and Clinical Health Act (HITECH) and the Genetic Information Nondiscrimination Act (GINA).[12] The federal regulations delegate authority to Institutional Review Boards (IRB)[a] to review and monitor any research involving human subjects to protect individuals from harm and preserve their rights. IRBs are required to conduct continuing review of research at intervals and to observe the consent process when necessary.

As the volume of research grew in recent years, the number of IRBs in the United States also increased substantially. A previous study shows that the number of IRBs has increased by 41% from 2004 to 2008.[13] According to the latest data from the OHRP, there are 2937 Institutional Review Board Organizations (IORGs), which include 3589 IRBs actively registered.

1.1 Problem Analysis

Clinical research professionals spend a great deal of their time preparing IRB applications. After a study is approved and open, amendments, application renewals, and unanticipated problems reports must be submitted to the IRB for continued monitoring.

[a] "IRB" is a generic term used by HHS and FDA to refer to a group whose function is to review research to assure the protection of the rights and welfare of the human subjects. Each institution may use different names such as Research Ethics Committee, Committee on Human Studies, the Committee on Clinical Investigations, etc. For the sake of simplicity, this dissertation will use the term IRB consistently.

The clinical research landscape has changed dramatically in recent years in terms of both volume and complexity. According to the latest data from ClinicalTrials.gov, the number of registered studies has increased 66% from 2010 to 2013.[14] Retrospective research involving secondary use of patient data has also been greatly encouraged and facilitated by the wide adoption of electronic health record (EHR) systems, which will be discussed in detail in Chapter 2. Recent years have seen a growing trend towards multisite clinical research, which offers numerous scientific advantages over single-site studies.[15,16] This changing nature of clinical research has caused new challenges for both clinical researchers and IRB reviewers. The remarkable growth of clinical research has resulted in an unprecedented increase in workload for IRBs, especially for large academic research institutions.[13] Variations in submission requirements, application formats, application questions, and review procedures across different IRBs have posed significant challenges specially for multisite studies.[17–19] Studies show that investigators complain that the IRB application process is burdensome and, in some instances, waiting to obtain IRB approval has delayed project initiation.[13,20–22] These issues are complicated by the political, educational and technical aspects of the human subject projection domain. A complete solution may include reforms in policies, administrative operational procedures, and technology uses. This dissertation focuses on only the technical aspect of the potential solution and considers other aspects to be beyond the scope of the dissertation. Information technology has been proven to have enormous potential for improving the performance of organizations.[23] It has contributed to high productivity in many domains, including but not limited to finance, retail, education, government, airline, and healthcare.

In the human subject protection domain, many institutions are employing some form of information technology to support IRB application submissions, tracking, and reviews. The term e-IRB will be used to refer to these information systems. According to the 2011 metrics report from the Association for the Accreditation of Human Research Protection Programs, Inc. (AAHRPP), shown in Figure 1, more than 90% of institutions use a database to track IRB applications. However, only about half of the institutions use an online system for IRB application processing, but the trend has increased from 2009.

The AAHRPP metrics report gives us a general idea of the information technology employment in IORGs. However, the capability of these IRB systems may vary greatly across organizations. Some institutions may electronically store only the descriptive information of the IRB applications for tracking purposes but not the IRB applications themselves. Some institutions may scan IRB applications from hard copies and store them as images. Some institutions may accept word or PDF IRB applications via e-mail or Flash disk and store them in a document management system like SharePoint.[24] In addition to the format varieties, the content and structure of IRB applications are specified by each institution and can vary enormously. The IRB application can include large sections of free text, or structured data elements that can be interpretable by computers.

Problems with the varieties of format and content of IRB applications across institutions are summarized as follows, whereas more detailed analysis will be discussed in Chapter 2 and 3.

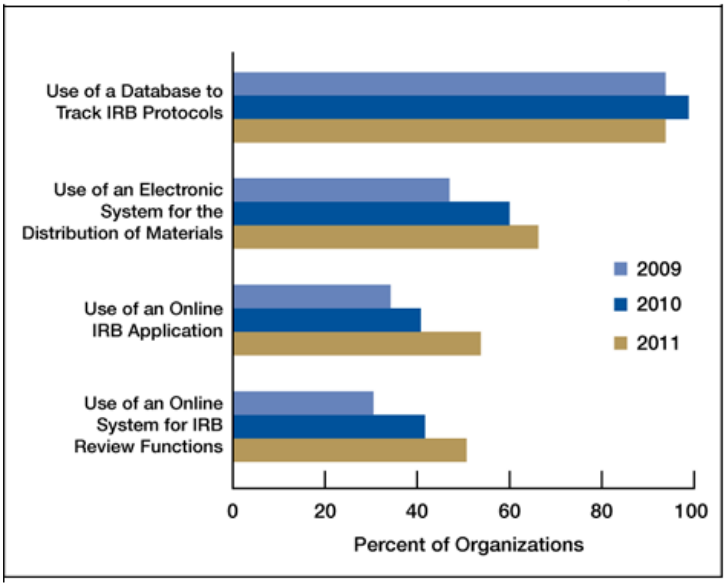

Figure 1. AAHRPP Metrics Report 2011: Technology Use by Organizations with an IRB Three-Year Trends

1. The content of the scanned paper IRB applications or IRB applications stored in word or PDF files is hard to search across documents. The IRB application repository is a valuable resource for archiving all human subject research conducted at an institution. It can be potentially helpful to different stakeholders such as investigators, reviewers, policy makers, and accreditation institutions for future analysis, performance measurement, reporting, and auditing. Investigators and IRB reviewers may need to use previously reviewed studies with similar features as a reference to help design and review a current study. For example, a commentary has proposed to organize information about incentives offered to participants with reference to a commonly recognized set of ethically salient parameters (e.g., condition under study, study population, types and levels of risks, research setting, etc.).[25] Currently, there are considerable variations in the types and amounts of incentives even for studies with similar features of protocol design and research setting. Information buried in scanned

images or unstructured documents will limit the ability to search and analyze such information across studies efficiently, thus hindering IRBs from developing and applying consistent incentive policies from comparable practices.

2. Free-text content is often ambiguous and lacks clarity in describing the study. This will impact the review quality and efficiency. For example, to ensure that study information, such as eligibility criteria and planned study procedures, is specified clearly and consistently throughout the protocol, IRB reviewers often need to ask for clarification from researchers. Free-text application information is hard for computers to process in order to assist human review. Alternatively, if the area of study is recorded in a more structured format, the application can be automatically assigned to IRB members with corresponding experiences and expertise. Depending on whether the planned study activities involve high-risk procedures such as ionizing radiation or only informational risk such as secondary use of existing data or specimens, commensurate review procedures for risk can be suggested by the system. This can help IRBs allocate their limited resources to ensure human subject protection while enabling responsible research to proceed.
3. The lack of uniformity of IRB application information impedes system interactions with other CRI tools, causing cumbersome workflows for clinical research. There is overlapping study information across different CRI systems. Without system integration, investigators need to enter duplicate study information into each system. For systems requiring IRB approval before certain actions can take place, the investigators may need to manually submit the IRB approval letter to different stakeholders. In addition, previous studies show that lack of standardized application

forms is one of the causes for inconsistency in ethics review for multisite studies.[26] The application format and content variations across different IRBs also make the application and review decision information sharing and exchange difficult, impeding efficient communication and cooperation among IRBs.

4. Each IRB expends effort developing local application templates or Web forms even though a large part of the content is essentially common across different IRBs in order to be compliant with best practices and regulations.

Despite the increasing interest and effort in developing and harmonizing standards for clinical research to share and reuse research data and metadata across projects and across institutions in the CRI domain,[27,28] e-IRB systems have not attracted much attention in the CRI community. However, IRB application and oversight is an important component in the clinical research process and can significantly affect the timeliness and ethical quality of the study.

1.2 Research Objective

To our knowledge, there is no existing research addressing e-IRB systems in the CRI domain. This dissertation aims to fill this research gap by:

1. Analyzing the IRB application systems used at 61 Clinical and Translational Science Award (CTSA) centers to understand the major application system types and gain a deeper insight into the diversities of IRB application systems within and across institutions.
2. Developing an IRB domain analysis model (IRB DAM) to support standard, implementation-independent, and machine-understandable representation of IRB application information in order to: a) support data interchange message development

between e-IRB systems and other CRI systems to streamline the clinical research process; and b) support a consistent and structured e-IRB design that makes automated review decision support possible, thus enhancing the quality and efficiency of ethics review.

3. Evaluating the static aspect of the IRB DAM by mapping five real-world IRB application systems with different designs from different institutions to the IRB DAM model.
4. Demonstrating a prototype implementation of the model by integrating an e-IRB system with a health data query system for secondary use to support automated access control on Protected Health Information (PHI) based on IRB review decisions.

The dissertation is organized as follows: Chapter 2 describes the original motivation of this work and related previous research; Chapters 3, 4, 5, and 6 describe the methods and results for each of the aims mentioned above. Chapter 7 discusses the limitations of the model and future work, and concludes the major contribution of the dissertation.

.

CHAPTER 2

BACKGROUND

2.1 Motivation

This research project was initially motivated by the need for automated PHI access control for the Federated Utah Research and Translational Health e-Repository (FURTHeR) project. FURTHeR is hosted in the Center for Clinical and Translational Science (CCTS) at the University of Utah and utilizes real-time data model and terminology translation services to transform an investigator's query to query representations appropriate for heterogeneous data sources and return federated results.[29] The current release of FURTHeR adapted the existing i2b2 (Informatics for Integrating Biology and the Bedside)[30] Web client front-end as the query user interface, allowing investigators to compose and execute queries and view aggregated, federated results.[31] However, to deliver individual-level PHI to investigators, appropriate and efficient security mechanisms must be employed to ensure patient privacy. This is a common challenge faced by most secondary use of health data systems, which will be discussed in further detail in the following subsections.

2.1.1 Secondary Use of Health Data for Research

Widespread adoption of information technology such as EHR systems in healthcare has driven tremendous growth of data and created greater opportunities for secondary use

of such data. Secondary use of health data for research purposes includes but is not limited to postmarketing safety surveillance of drugs and devices, health services research, comparative effectiveness analysis, and identification of potential participants for clinical trials.[32–35] Secondary use of health data is one of the most promising ways to improve patient care with lower cost and greater timeliness.

Efforts are increasing towards building informatics tools to support secondary use of health data for research purposes. The term SUHD will be used to refer to this type of information system. Many hospitals have established enterprise data warehouses (EDW) or disease registries that extract patient data from the operational EHR environment into a central repository designed to support research and quality improvement projects. There are also cross-institutional and national efforts towards developing informatics infrastructures and standards to facilitate access to multiple data sources, including but not limited to the cancer Biomedical Informatics Grid (caBIG®),[36] i2b2,[37] FURTHeR, the Query Health Initiative,[38] and the SHARPn platform developed by the Strategic Health IT Advanced Research Projects Area 4 Consortium.[39]

Some SUHD systems provide a query interface directly accessible to researchers to enhance clinical research efficiency. The i2b2 Web client is one of those tools that has been popularly adopted or adapted. It allows users to build and run queries that specify clinical criteria for a desired cohort by using simple drag-and-drop methods in a Web browser client. The Shared Health Research Information Network (SHRINE), for example, deployed the i2b2 software suite in a federated environment and can return aggregated counts of eligible subjects for preparatory-to-research activities.[40] The Cohort Discovery tool, deployed at the Oregon Clinical and Translational Research

Institute, also utilizes i2b2 to allow users to directly discover patient cohorts. Patient sets can be extracted from queries only with proper IRB approval.[41] The STRIDE (Stanford Translational Research Integrated Database Environment) Cohort Discovery Tool (CDT) from the Stanford Center for Clinical Informatics allows Stanford researchers to query the STRIDE Clinical Data Warehouse (CDW) and returns the approximate number of patients matching the search criteria.[42] No patient identifiers or clinical data, however, are revealed before Stanford IRB approval is obtained. The Query Health Initiative led by the Office of the National Coordinator for Health Information Technology (ONC) is designed based on the principle of "bringing questions to data" instead of the conventional approach "bringing data to questions," in which data sources ultimately retain control over the decisions whether to respond to a distributed query as well as maintaining control over the data to be released. The current Query Health reference implementation supports only aggregated query results.[43]

2.1.2 Privacy Concerns

A secondary use healthcare data repository and a query interface directly accessible to researchers provide a more convenient approach for preparatory to research activities. However, as demonstrated above, most of the existing query interfaces allow aggregated results only for data requests and none of them allow streamlined individual-level data access without manually checking for IRB approval due to privacy concerns.

There are efforts towards de-identification of health data for research purposes since de-identified data are not covered by the HIPAA Privacy Rule. However, de-identification cannot completely solve privacy issues for the following reasons. First, de-identified data are not sufficient for many studies that require PHI for analysis. In other cases,

researchers may need PHI to recruit participants. Second, there are re-identification risks for HIPAA Safe Harbor and Limited Dataset policies. It is difficult to fully de-identify health data. Third, even though de-identified data are not regulated under the HIPAA Privacy Rule, there are local or institutional policies for accessing de-identified data, which usually require reviews from IRBs, although this could be an exempt or expedited review instead of a full review.

Therefore, one of the major concerns raised by these SUHD systems is how to handle data requests from researchers and deliver the desired data set accurately and efficiently without compromising patient privacy. While maintaining privacy and confidentiality is essential to ethical research, insufficient oversight procedures can also hinder progress in biomedical research and create cumbersome research processes. Problem with the current IRB oversight of SUHD studies were illustrated in detail in our previously published paper,[44] from the point of view of the investigator, the data provider, and the IRB.

Although there are discussions in the advance notice of proposed rulemaking released by HHS about whether data-only research projects posing informational risks (e.g., resulting from the unauthorized release of information about subjects) should be reviewed by IRBs,[45] this is considered as a regulatory issue that is outside the scope of this research project. Our goal is to employ information technology to streamline the research process while staying compliant with current regulations.

2.1.3 Automated Access Control of PHI Based on IRB Approval

The inefficient and time-consuming process for accessing PHI is caused by the information disconnect between SUHD systems and e-IRB systems. In a previous paper, we proposed a framework that integrates FURTHeR with the University of Utah's e-IRB

system, ERICA, to achieve automated access control on PHI based on IRB approval.[44] The integrated framework aims to streamline the current data request and access workflow for researchers, IRBs, and data custodians. The data elements requested by an investigator and associated IRB review decisions will be shared between the two systems in a meaningful way so that the IRB knows what data elements are requested by the investigator and the data query system knows what data elements are approved by the IRB for an investigator to access.

2.2 Why a Model Is Needed

To enable meaningful information exchange between FURTHeR and ERICA as discussed in the use case scenario in section 2.1, a clearly defined message structure and content representing data query criteria, data access requests and IRB review decisions need to be specified. Considering that FURTHeR is designed to connect to multiple data sources across institutions, IRB review from multiple institutions may be needed. In addition, since SUHD systems emerging at other institutions have similar access control needs, it is valuable to establish a shared view of the information exchanged between any SUHD system and any e-IRB system.

However, SUHD systems are only one of the many types of CRI systems and there are potential benefits in integrating e-IRB systems with other CRI systems to streamline the clinical research process. For example, by integrating CTMSs with e-IRB systems, study protocol information can be shared and researchers do not need to enter duplicate information twice or many times when IRB applications to multiple sites are required. Any subsequent protocol change updated in the CTMS can be transmitted to e-IRB system(s) in a timely and accurate fashion. Study status reports required for IRB

continuing review can also be automatically generated in CTMS by specifying corresponding parameters in the system. EDC systems, either implemented as part of CTMS or as a standalone system, by integrating with e-IRB systems, can ensure that no data can be collected in a study before IRB approval. Unanticipated adverse events can also be transmitted to the e-IRB systems from EDC or CTMS via a direct machine-to-machine connection. For multisite studies, sometimes one IRB needs to rely on other IRBs' review decisions. Integration between different e-IRB systems can enable application information and review decision sharing, thus making the communication more efficient between different IRBs and eliminating the needs for investigators to submit multiple applications for the same study protocol. Therefore, data sharing and exchange between e-IRB systems and other CRI systems can greatly streamline the clinical research process.

Information models are the basis for retaining computable meaning by refining message structure and content when data are exchanged between heterogeneous computer systems. Without a standard information model for e-IRB systems, point-to-point system integrations will be needed that can cause a large amount of interface development. The model should be designed to represent information for the IRB domain in a higher level of abstraction in order to be generalizable across institutions and systems. This is considered a domain analysis model (DAM) in software engineering. With this IRB DAM:

1. Existing e-IRB systems can use this model to develop a standard interface layer to achieve semantic interoperability with other CRI systems to streamline the clinical research workflow. For multisite studies, sharing application information among IRBs

in participating institutions can facilitate joint review or federated review by relieving duplicative burdens on both IRBs and investigators.

2. Structured information in e-IRB systems based on the model can enhance review quality and efficiency via automated review decision support enabled by utilizing the structured study protocol information and defined rules based on federal and local regulations. Computer-assisted review support is not intended to replace human review but to promote a partnership between human and computer-based resources and capabilities so the strengths of human beings and computers can be fully utilized. In addition, with structured IRB application information, it is easy to search across studies. An IRB considering a new protocol could search the application database and find similar studies and related ethical concerns and review decisions, which can be used as a reference for current review.
3. A structured IRB application repository can become a valuable resource for various stakeholders in the research domain. This can make secondary use of human research metadata projects such as the Human Studies Database (HSDB) Project[46] much easier. It can also be useful for developing practical guidance for effective IRB oversight of human subject protection and developing an ongoing process to increase empirical knowledge about current “best practices.”
4. New e-IRB systems can be developed using the IRB DAM as a reference model, thus avoiding the need to start from scratch. It will also enhance the compliance with federal regulations since the model covers major federal regulations. Even though an institution does not have the resources to develop or buy an e-IRB system at the

moment, the model can be used as a reference for designing application forms in word processing templates.

2.3 Related Work

To the best of our knowledge, there is no existing research on modeling the IRB oversight domain. However, there have been a number of efforts to model various aspects of biomedical research in general.

2.3.1 CDISC Standards

The Clinical Data Interchange Standards Consortium (CDISC) has developed a series of platform-independent standards to support the acquisition, exchange, submission, and archiving of clinical research data and metadata. The most relevant model from CDISC is the Protocol Representation Model (PRM), which is a representation of the elements of study protocol. It focuses on the characteristics of a study and the definition and association of activities within clinical trial protocols, including “arms” and “epochs.” It is designed to facilitate protocol information to be reused and repurposed across multiple documents, databases, and systems from study start-up through reporting and regulatory submissions.[47]

2.3.2 BRIDG

The Biomedical Research Integrated Domain Group (BRIDG) project developed a comprehensive domain analysis model for protocol driven research and its associated regulatory artifacts by coordinating the standards development efforts in HL7 Regulated Clinical Research Information Management (RCRIM), CDISC, and the National Cancer Institute (NCI).[48] The goal of the BRIDG project is to develop computable semantic

interoperability by standardizing and exchanging clinical trials data among different projects. BRIDG defined six subdomains: Adverse Event, Common, Protocol Representation, Regulatory, Statistical Analysis, and Study Conduct. It covers concept entities defined for planned, scheduled, and performed clinical trial activities. Certain subdomains in BRIDG such as Study Conduct, which includes the detailed scheduling and conduct of study activities are designed for CTMS systems and are not of interest to IRB. The Protocol Representation subdomain is developed based on PRM from CDICS, as mentioned above. It defines concepts and attributes associated with prospective clinical trials but lacks structural attributes related to retrospective studies, which compose a major category of human subject research. The Regulatory subdomain was originally designed for FDA regulated product submissions instead of IRB oversight. It is also too coarse-grained for computers to process by modeling the submissions at a document level.

2.3.3 The Ontology of Clinical Research (OCRe)

The Ontology of Clinical Research (OCRe) is a formal ontology for annotating human studies and supporting federated query on data and meta-data across various studies. Compared to BRIDG, which covers all phases of a clinical study, OCRe focuses on the design and analysis phase of studies. The *study_protocol* module in OCRe incorporates related classes from the BRIDG model. OCRe is designed as an OWL ontology that models the entities and relationships of human studies and it has been used only for annotating existing study metadata (e.g., study design, eligibility criteria, interventions and observations, etc.) from different sources to support federated query of

studies. [46] OCRe is not designed for data sharing. However, it may later be used as a terminology binding for certain values sets in our IRB model.

2.3.4 The Ontology for Biomedical Investigations (OBI)

The OBI project is an international, collaborative effort to build an ontology to be used for annotation of Biomedical Investigations. The origin of OBI is in the molecular biology research domain.[49] Although it is aimed at developing cross disciplinary, integrated ontology including a set of broadly applicable terms for the detailed description of biological and clinical investigations, the current release still focuses mostly on terms describing biological experimental processes and their relevant components (e.g., Antigen binding, T cell activation, Immunization in vivo, etc.) It also lacks the definitions for the structure of a study protocol. Specifically, it defines the class *protocol*, which has only a subclass *animal care protocol*, which in turn has a subclass *rodent care protocol*. The *protocol* class has only two "has part" relations with *objective specification* and *action specification*, respectively.

2.3.5 Biomedical Ethics Ontology

Koepsell et al. proposed to develop an ontology for the biomedical ethics domain. Their paper described a vision of using the Biomedical Ethics Ontology (BMEO) to support automated informed consent document generation, sharing review opinions and decisions across ethics committees, automated review decision support, and reducing the vagueness in the review process by formalizing rules and regulations.[50] The ontology will include rigorous definitions of biomedical ethics terms such as *autonomy, informed consent, minimal risk, harm, study, research, human subject*, etc., and their relations. A

follow-up publication by DuBois questioned Koepsell's paper by pointing out that ontology is ill-suited to describe regulatory definitions and ethical concepts that are not universal.[51]

Although one of the goals of our project overlaps with the BMEO project (facilitating automated review decision support), I took a different approach and design principle. Our principle is to first capture all necessary information pertinent to a human subject research that is essential for IRBs to evaluate the study. Such information will be captured and represented in a structured way whenever possible, which is the foundation for later review decision support based on predefined rules or through binding to an ontology. I try to avoid modeling verbatim concepts defined in regulations unless they have a clear extension definition (e.g., vulnerable population is defined by federal regulations and OHRP guidelines via naming a list of vulnerable subject categories[52]). I do not model entities that do not require computer interpretation such as such as the standard language used in an informed consent, although it is possible to facilitate automatic informed consent generation in the future as a more advanced application of the model. Our goal is not to replace human review with a computer system, but to make the reviewer's job easier by utilizing information collected in a machine-understandable way. In this way, I am trying to maximize the advantages of both human (IRB reviewers) and automated processes, while adding the capability to integrate IRB system into other CRI applications.

CHAPTER 3

ANALYSIS OF IRB APPLICATION SYSTEMS AT CTSA INSTITUTIONS

The AAHRPP metrics on technology use of IRB application systems gave us a general idea about how applications are managed among U.S. IORGs. However, it is not clear what specific types of "online systems" exist. If we build a model, will our model be applied or useful to other institutions? To gain a deeper insight into the IRB application systems currently used in the U.S., I did an analysis of the IRB application systems currently used at CTSA institutions.

The reason I chose CTSA institutions as the sampling target for the analysis is 1) CTSA institutions are leading academic medical centers and large healthcare delivery organizations in the U.S. and they are research intensive. A previous study shows that IRBs at large institutions have a heavier workload compared to small institutions;[13] 2) The CTSA program aims to build an environment to increase the efficiency and speed of clinical and translational research. CTSA institutions have a strong focus on adoption of CRI systems to achieve this goal. For example, according to the results from a 2010 CTSA survey, 86% of CTSA institutions have an integrated data repository (IDR).[53] Many institutions are actively deploying electronic data capture tools such as REDCap or CTMS such as OnCore[54] to support clinical and translational research.[55],[56]

Therefore, the potential for system interaction between these CRI systems and e-IRBs can greatly benefit CTSA institutions by streamlining the research workflow and improving research efficiency.

3.1 Analysis Methods

All CTSA centers are listed at the CTSA consortium Web site (https://www.ctsacentral.org/institutions). I did the analysis by exploring the official IRB websites at each CTSA center and its participating institutions. Most IRB websites have a detailed IRB submission guideline or training documents or videos for investigators, from which I can directly tell what type of submission system is used and the system's vendor if it is commercial software. For those institutions that do not have such information on their websites, I called or emailed the IRB office with the following questions:

1. How does an investigator submit IRB applications? Hard copy, E-mail, portable device, or online?
2. If an online submission system is used, is it developed in-house or purchased from a vendor?
3. If it is a commercial software product, which vendor is it?

If I called the IRB office more than three times and still could not reach the relevant personnel, then I counted it as unknown and excluded it from the final analysis. Compared to the regular survey or interview approach, our approach-- making full use of online resources-- avoided problems such as low response rate and long waiting periods that are common challenges faced by survey studies. At the same time, it gave us the same quality of insight into the IRB systems selected.

3.2 Analysis Results

All 61 CTSA centers were included in the analysis. Each CTSA center may have one or more partner institutions. For example, the University of Utah Center for Clinical and Translational Science (CCTS) has three participating institutions: University Health Care, Intermountain Health Care, and Veterans Administration Salt Lake City Health Care System (VA SLC). Another CTSA center, the Harvard Clinical and Translational Science Center, is composed of 32 partner institutions. For those CTSA centers that have more than one partner institution, I chose two to three to analyze their IRB application systems. I ended up analyzing 121 individual institutions in the 61 CTSA centers. Among the 121 individual institutions, there are 103 IORGs in total. This is because not every institution has their own IORG but some of them rely on an external IORG from a partner institution. For example, among the three partner institutions in the University of Utah CCTS, University of Utah and Intermountain Healthcare have their own IORGs, whereas VA SLC uses the University of Utah's IORG. Usually one IORG uses a single submission method across different IRB panels but there are also exceptions. Among the 103 IORGs I analyzed, there is one IORG using two different submission systems for different IRB panels. Therefore 104 submission systems were analyzed in total. Eight of the 104 systems were excluded because I could neither find relevant information on their website nor reach the relevant personnel by phone or email. The 96 systems are categorized into three top categories: Hard Copy, E-mail (Flash Drive or CD), and Online Submission as shown in Table 1.

Online submission systems were further categorized into five subcategories:

1. Commercial e-IRBs: e-IRB systems that were purchased from vendors. Examples of

Table 1. Analysis Results of IRB Application Systems Used in CTSA Institutions

System Type		Number of Systems	Percentage
Hard Copy		10	10.4%
E-mail (or Flash Drive, CD)		17	17.7%
Online		69	71.9%
	Commercial e-IRB	31	32.3%
	In-House e-IRB	19	19.8%
	SaaS e-IRB	6	6.2%
	Community Source e-IRB	3	3.1%
	Not specified	10	10.4%
Total		96	100%

such e-IRBs are iRIS™ by iMedRIS™, Click IRB by Huron Consulting Group (originally Click Commerce), InfoEd by InfoEd Global, etc.

2. In-House developed e-IRBs: e-IRB systems that were developed internally by the institution's technology team.
3. Software as a service (SaaS) e-IRBs: The e-IRB system and submitted IRB applications were hosted by a vendor and made available to customers as services. SaaS e-IRBs are accessed by users using a Web browser. Examples of SaaS e-IRBs are IRBNet, IRBManager, etc.
4. Community source e-IRBs: A special type of open source software with institutional contributions by colleges, universities, and some commercial firms rather than from

individuals. Kuali Coeus[57] is an example of a community source e-IRB.

5. Other online e-IRBs with unspecified system types.

About 20% of IORGs use independent commercial IRB Services such as Western IRB and RCRC IRB as a complementary IRB review method for industry-sponsored clinical studies.

3.3 Discussion

From the analysis, I obtained a deeper insight into the IRB application systems used across the nation. Our major findings are listed below:

1. As shown in Table 1, there are 71.9% IORGs using some type of online submission, which is higher than the AAHRPP metrics report sampled across the nation.
2. Online submission is not equal to structured IRB application information. Some online submission systems simply allow investigators to upload the application document in Word or PDF format. Even for those systems that support Web forms, there is still a large proportion of free-text fields. Unstructured information is difficult to be processed by computers for further analysis or data sharing.
3. The CTSA institutions prefer to purchase commercial software or develop their own e-IRB systems to SaaS e-IRBs, which are more suitable for small hospitals with limited resource.
4. The design of each institution's application templates or forms varies. Most commercial e-IRB systems such as iRIS™ and Click IRB support customizable form, which allows flexible localization. However, this also means that two institutions using the same e-IRB product may have totally different application form designs.

The varieties of IRB application forms will impede the interconnection between e-IRB systems and other CRI tools.

5. Many of the commercial e-IRB systems used by CTSA centers are part of the whole Electronic Research Administration (eRA) solution from vendors. A single vendor may provide a suite of software products for all kinds of research administration activities such as grant proposal preparation and submission, award management, publication management, etc. Besides the eRA solutions, some vendors also have software products for the actual conducting of a study, such as CTMS or EDC systems. Similar to the discussion about best-of-breed or single vendor Electronic Health Record (EHR) systems in the healthcare domain, there exists the same pattern in the clinical research domain—using a research package from a single vendor or using best-of-breed modules from multiple vendors. A single vendor solution has the advantage of better integration between different systems or modules, which provides a more streamlined workflow. However, a single vendor solution may not meet all the needs of an institution or the cost may be too high for a small institution. A detailed discussion about this issue is beyond the scope of this dissertation. The point is for the best-of-breed CRI solution, systems from different vendors or developed in-house need to be integrated to achieve a streamlined research workflow. Even for the single vendor solution, system integration is needed when multisite studies are conducted at multiple institutions.
6. The IRB application systems used in one institution can be very diverse. One university can have multiple IORGs for different departments. Each of them may use a different IRB application system.

CHAPTER 4

MODEL DEVELOPMENT

Chapter 2 described the motivation of this dissertation project. The need for developing a standard information model for e-IRB systems in order to promote system interoperability in the clinical research domain was identified. In Chapter 3, the variations of IRB application systems were further analyzed. Considering a higher percentage of e-IRB deployment and the support for customizable form design by some e-IRB systems, the potential for existing IRB application systems in large academic institutions to adopt a standard model was demonstrated. This chapter describes the method of the IRB domain analysis model development, which is primarily based on existing domain analysis and data modeling methodologies in software engineering and modeling efforts in the healthcare and research domain. The result of the IRB DAM includes the static (structural) entity-relationship model, the dynamic (behavioral) business process model, and the interaction model. A preliminary terminology binding effort to develop a domain vocabulary specification for data request related attributes is also presented in this chapter.

4.1 History of Domain Analysis

Before I describe the details of the IRB domain analysis model, I should define what a domain analysis model is. The term *domain analysis* was first introduced in software engineering by Neighbors to describe "the activity of identifying the objects and

operations of a class of similar systems in a particular problem domain."[58] In domain analysis, common characteristics from similar systems in an application domain are generalized. It is thus at a higher level of abstraction than *system analysis*, which focuses on a specific system. Since then, domain analysis (DA) attracted lots of attention and has been considered a key factor in successfully creating reusable components in software engineering by reuse of analysis and design information, in addition to programming language code.[59,60]

A host of documented domain analysis methods are available in software engineering and there is no single standard domain analysis process defined. Ferre et al. evaluated the various DA methods and did a comparison of each method by the type of artifacts to be reused.[61] Different DA methods have different goals, and the DA process or representation for a domain analysis model varies according to the goals to be achieved. There is little agreement on the definition of the term *domain analysis* since the author's definition is strongly influenced by the expected outcome of the DA process.

On the other hand, some researchers may use a different term such as "*Conceptual Modeling*" to represent an abstraction modeling process similar to domain analysis.[62] Conceptual models are designed to describe the semantics of software applications at a high level of abstraction, which enables all stakeholders, including domain experts, analysts, and application programmers, to understand and communicate with each other successfully.

HL7 defines a Domain Analysis Model (DAM) as "an abstract representation of a subject area of interest, complete enough to allow instantiation of all necessary concrete classes needed to develop child design artifacts." [63] It emphasizes that a complete

DAM includes both static (e.g., class and instance diagrams) and dynamic (e.g., activity and state diagrams) semantics of a domain. The HL7 Healthcare Development Framework (HDF) adapted a popular DA method in software engineering—Feature-Oriented Domain Analysis (FODA)—to the healthcare business. FODA was developed by the Carnegie Mellon Software Engineering Institute (SEI) and focuses on the identification of distinctive features of software systems within a domain.[64]

4.2 Domain Analysis Methodology for the IRB Model

Domain analysis methods such as FODA in software engineering are mostly designed for software reuse. The conceptual modeling method presented by Embley et al. promotes that conceptual models can go beyond being mere blueprints; rather, they can constitute the basis for automatically or semiautomatically generating the software system itself. Although our IRB DAM has potential values in software reuse and model-driven software development, the major goal of the IRB DAM is to achieve interoperability among systems in the clinical research domain. Therefore, I adapted the FODA method and the conceptual modeling method in order to meet this requirement. Figure 2 illustrates the overview of the IRB domain analysis process and the corresponding model artifact generated by each step.

The first phase, context analysis, defines the scope of the modeling domain. In this phase, the relationships between the candidate domain and the elements external to the domain are analyzed. The second phase, domain modeling, consists of entity-relationship modeling, which models the static (structural) semantics of the domain, and business process modeling, which models the dynamic (behavioral) semantics of the domain. The static artifact produced from entity-relationship modeling is called the entity- relationship

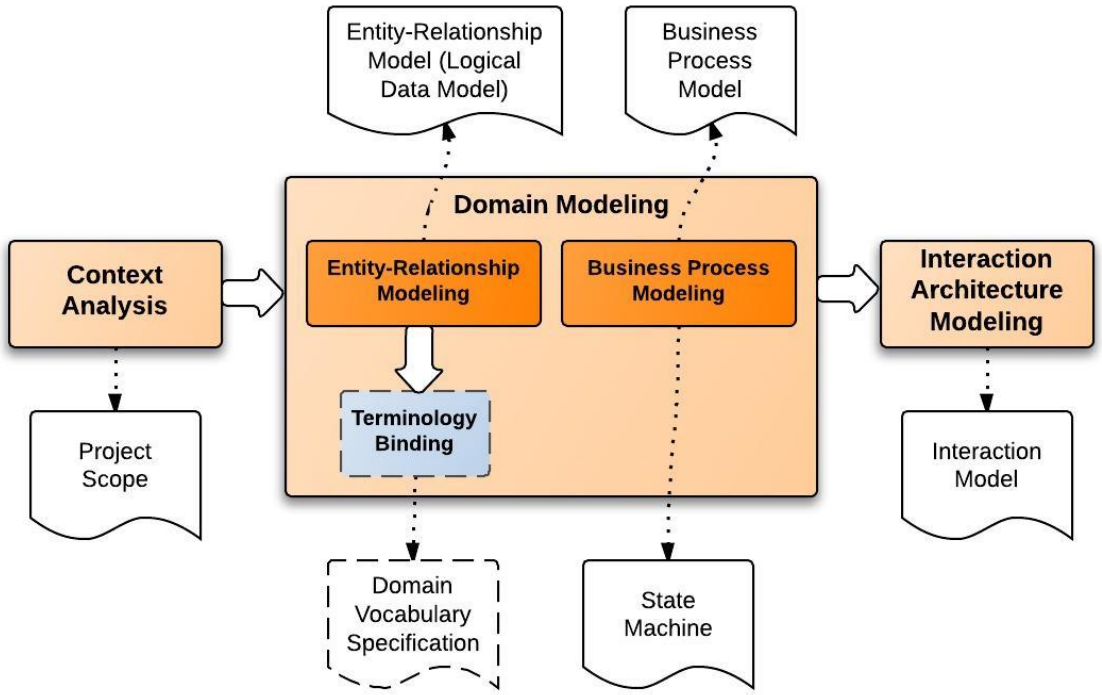

Figure 2. Overview of the IRB Domain Analysis Process

model. It is independent of the underlying database design. Some literature refers to this as an information model or logical data model. In this dissertation, these terms will be used interchangeably. Terminology binding connects the entity-relationship model with domain vocabulary specifications to standardize the terminology that describes the domain. In this dissertation, only a preliminary terminology binding effort was conducted. A comprehensive domain vocabulary development is considered as a future development. The interaction architecture modeling phase specifies interactions and information exchange between an IRB system and other clinical research information systems to realize new features that are not supported by isolated systems.

The entity-relationship model, the business process model, and the interaction model are formally represented using the Unified Modeling Language (UML). The following subsections will describe each phase of IRB domain analysis process in detail.

4.2.1 Context Analysis: Domain Scope

The modeling scope of the IRB oversight domain is defined during the context analysis phase. Figure 3 illustrates the scope of the modeling domain and its relationships with its parent domain, subdomains, and peer domains.

The IRB oversight domain is a subdomain of human subject research, which is a subdomain of the general biomedical research domain. During our analysis of CTSA IRB systems, I found that although federal regulations mandate only that IRBs review human subject research, it is not uncommon for IRBs to require investigators to submit an application for studies considered to be nonhuman subject research (e.g., studies using de-identified data for secondary analysis) and the IRB will make the final decision whether the study is nonhuman subject research. Our model does not reflect this type of study, since this is determined by institutional policy and it is considered to be an implementation-level issue. However, the model is extendable to support this type of study in an IRB system.

Human subject research has several components throughout the life cycle of the study. IRB oversees only certain aspects that are closely related to protection of human subjects. There is significant overlap within the IRB and the study protocol domain. Study protocol is the blueprint of every human subject research project and it gives the IRB a comprehensive view of the study, including subject eligibility criteria and recruitment, planned study procedures and interactions, data management, and analysis plans, etc. It is essential for IRBs to evaluate the potential benefits and risks of a study. There are also certain aspects of a study that might not be covered in the study protocol but are important for IRBs to evaluate such as the informed consent process, compensation to

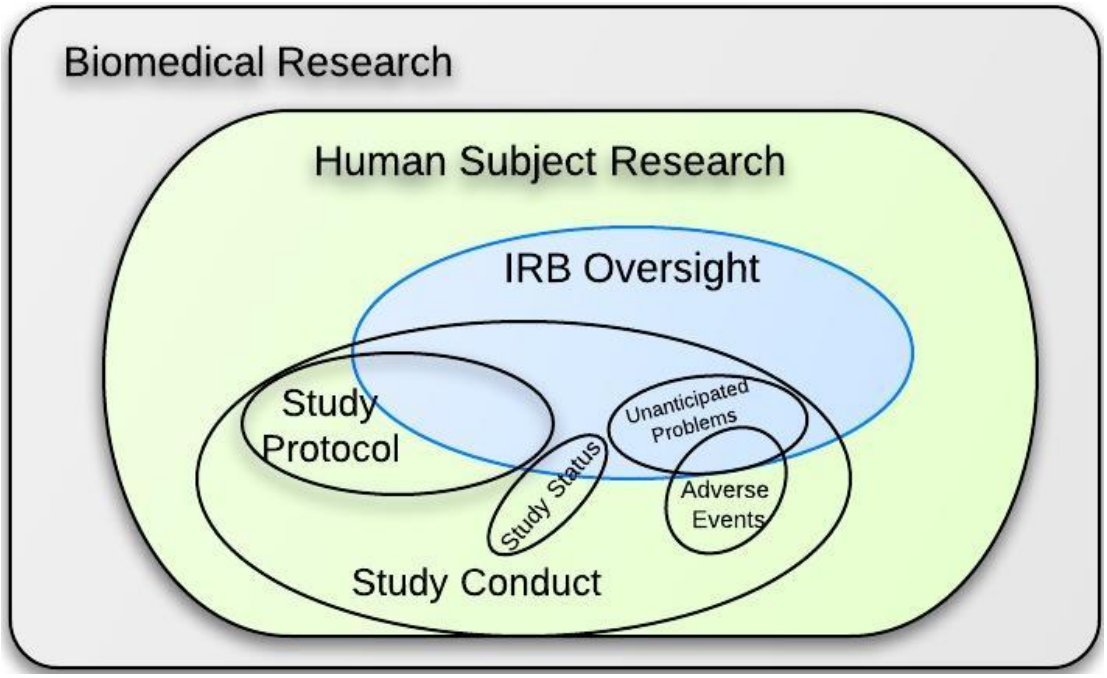

Figure 3. Context Analysis for the IRB Oversight Domain

subjects, vulnerable population participants, etc.

IRBs are also responsible for monitoring the conduct of human subject research. Adverse event (AE) tracking is a subdomain of human subject research. However, not all AEs are reportable to IRB. According to guidance from the Office for Human Research Protections, only unexpected AEs that are related or possibly related to participation in the research and cause a greater risk of harm than previously known should be reported to IRBs.[65] There are also other types of reportable unanticipated problems that are not AEs, such as a data breach or product (e.g., drug or medical device) problem. In addition, status reports on ongoing studies should be submitted to the IRB as part of application renewal.

4.2.2 Domain Modeling

The structural entity-relationship model is the core of the IRB DAM. It is used to acquire knowledge about the domain by modeling all physical and informational entities in the IRB oversight domain (e.g., study protocol, unanticipated problem, oversight committee, review record, etc.) and their relationships. It is the basis for supporting any internal or external automated tasks. The entity-relationship model is the most important artifact of the IRB DAM and will be described in detail.

The behavioral business process model specifies the high-level features of an e-IRB system, with a focus on end-user perspective of the functionality of the application. Considering the potential variability in review workflows among different IRBs, the current business process modeling intentionally avoids detailed workflow design such as application review assignment, internal review processes, or meeting scheduling. System interactions with clinical researchers and IRB reviewers are modeled and considered to be applicable across institutions. The state transitions of an IRB application are also represented as part of the business process model.

4.2.2.1 Modeling Tool

The IRB DAM is represented in Unified Modeling Language (UML) and developed using Enterprise Architect (Version 9.2) from SPARX Systems. UML is a standard general-purpose modeling language that can be applied to all application domains (e.g., health, finance, telecom, aerospace, etc.).[66,67] It was developed in an effort to consolidate the large number of object-oriented development methods in software engineering that had emerged and it is now maintained by the Object Management Group, Inc. (OMG). Traditionally, UML has been associated with object-oriented software

engineering and system design. However, with the rich modeling capabilities of UML 2.x, UML has been commonly employed in modeling application domains in addition to specific systems.

UML is a visual modeling language that can represent different perspectives of a system or a domain in *diagrams*. A diagram is the graphical presentation of a set of elements, most often rendered as a connected graph of vertices and arcs. There are many types of diagrams in UML and only a few of them were used for the IRB DAM. The UML Class Diagram is used to represent the entity-relationship model. The Business Process Model and Notation (BPMN) Process Diagram is used to represent the business process models. The UML State Machine Diagram is used to describe state transitions of an IRB application.

4.2.2.2 Information Sources

The FODA recommended several types of information sources to use while gathering information for domain analysis, including textbooks, standards, existing applications, and domain experts. Considering the special regulatory characteristic of the IRB domain, the content of the IRB model was driven by analysis of the following knowledge sources:

4.2.2.2.1 AAHRPP Accreditation Standards

The Association for the Accreditation of Human Research Protection Programs, Inc. (AAHRPP) is an independent, nonprofit organization that accredits high-quality human research protection programs in organizations in the United States as well as foreign organizations that are obligated to follow U.S. federal regulations.[68] The goal of the AAHRPP is to promote ethically sound research and ensure that all human research

participants are respected and are protected from unnecessary harm. It has published a series of accreditation standards that are compliant with regulations from the U.S. federal agencies that oversee research with human participants such as the HHS and the FDA, as well as other departments or agencies that have additional requirements, such as the Department of Defense and the Department of Veteran Affairs. The accreditation standards also cite international ethical principles such as the Declaration of Helsinki from the World Medical Association (WMA) and the Good Clinical Practice Guideline (GCP) from the International Committee on Harmonisation (ICH).

Specifically, the Evaluation Instrument for Accreditation (Version January, 2012),[69] and IRB Evaluation Checklist (Version December, 2010)[70] were used as guidelines in the IRB domain modeling process. The AAHRPP's Evaluation Instrument for Accreditation defines three *Domains* of responsibility: Organization, IRB or Ethics Committee (EC), and Researchers. The IRB modeling used the "IRB or Ethics Committee" domain. Within each *Domain* is a list of *Standards*, and for each *Standard* there is a list of *Elements* that provide more specificity. A total of 5 *Standards* and 25 *Elements are* defined in the "IRB or Ethics Committee" domain. The original Code of Federal Regulations (CFR) that are cited by each *Element* were reviewed for clarification or validation where needed.

AAHRPP accreditation standards were used as a knowledge source because they incorporate major regulations and ethical principles on human subject protection, thus providing the domain analysis an easily accessible comprehensive knowledge source. The documents are well organized and written in an easy-to-understand language as an interpretation of the federal regulations. This makes the domain modeling much easier

compared to directly using regulations as a knowledge source. In addition, more and more organizations are seeking AAHRPP accreditation for their human subject protection program. The IRB entity-relationship model can be used as a reference model for IRBs to design their application forms. Although this does not guarantee complete AAHRPP compliance, since AAHRPP also requires various operational procedures that are not reflected in the static IRB model, collecting necessary study information is the basis for developing standard operational procedures that are recommended by AAHRPP.

4.2.2.2.2 Guidance Documents from OHRP and FDA

The Office for Human Research Protections (OHRP) and the FDA both publish a variety of policy and regulatory guidance materials to assist the research community in conducting ethical research that is compliant with the HHS regulations.[71] These guidance documents address various topics in detail pertaining to human subject protection, such as the informed consent process, vulnerable population protection, investigational use of drugs and medical devices, etc.

4.2.2.2.3 HIPAA Privacy Rule and Related Educational Materials

The AAHRPP accreditation standards do not include regulatory items related to IRB from the HIPAA Privacy Rule. Educational materials provided at the NIH Web site were used for the domain modeling since they are easy to understand. Specific documents include “Clinical Research and the HIPAA Privacy Rule,”[72] “Institutional Review Boards and the HIPAA Privacy Rule,”[73] “Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule,”[74] and “Research Repositories,

Databases, and the HIPAA Privacy Rule."[75] I referred to the original Privacy Rule whenever necessary.

4.2.2.2.4. Existing Standard Models in the Biomedical Research Domain

As discussed in the context analysis section, the IRB oversight domain is a sub-domain of the human subject research domain, which is a subdomain of the general biomedical research domain. The existing standard models in the biomedical research domain from BRIDG and CDISC PRM overlap with the IRB model in certain areas, such as study protocols and adverse events. Since BRIDG already includes CDISC PRM, only the BRIDG model is used as a knowledge source for the IRB domain modeling. Model reuse has the following benefits: First, as domain experts have already devoted time to defining terms in these models I save myself substantial effort by not replicating that work. Second, by considering standard models during the model development phase, I made it easier to integrate the IRB model with the BRIDG model in future harmonization processes. Third, by reusing existing models that others already use, we improve the potential for future system interoperability.

4.2.2.2.5 Domain Experts

Two IRB domain experts were interviewed during the model development process. They both serve as IRB committee members at the University of Utah and have extensive experience in reviewing IRB applications.

4.2.2.3 Domain Modeling Process

The IRB entity-relationship model is the core of the IRB DAM. Figure 4 illustrates the process for the entity-relationship modeling. The business process modeling employs

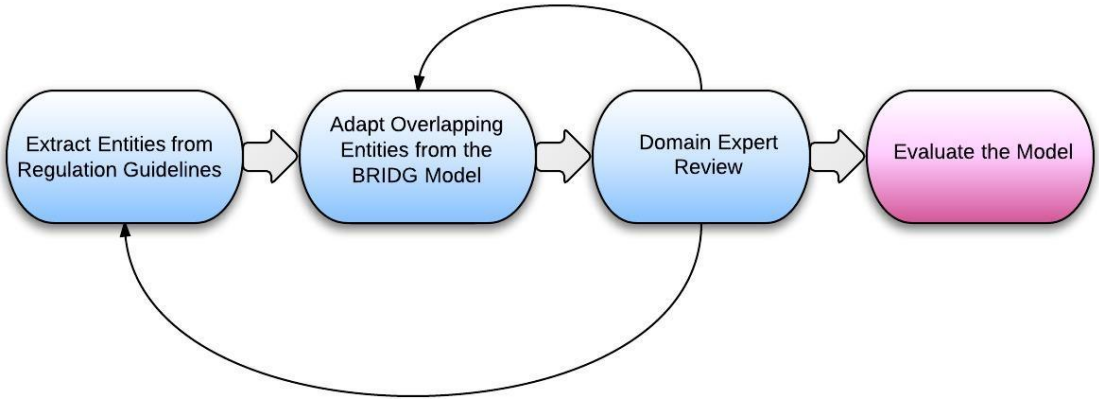

Figure 4. Overview of the Entity-Relationship Modeling Process

a simplified procedure, which primarily involves iterations of process analysis, design, and review.

First, all the key entities and their attributes in the IRB oversight domain were extracted from regulation-based guidelines including the AAHRPP accreditation standards, OHRP and FDA regulatory guidance documents, and HIPAA educational materials. For example, in the AAHRPP Evaluation Instrument document, *Element* II.2.D defines all the information elements that should be included in a status report for continuing review by a convened IRB. The model then defined the entity "*StatusReport*" with each information element as an attribute of this entity.

Second, relevant entities and their attributes from the BRIDG model were reused especially when the regulation guidelines did not specify the necessary details of the domain. AAHRPP accreditation standards do not always list the specific information that should be reviewed by the IRB. For example, in the same *Element* II.2.D, AAHRPP only states a general operational requirement about the information that should be provided to the IRB for initial review as follows:

> For initial review of research by a convened IRB, policies and procedures indicate that when they are scheduled to attend an IRB or EC meeting, all members (including attending alternate members) are provided and review: The full protocol, application, or a protocol summary containing the relevant information to determine whether the proposed research fulfills the criteria for approval.

AAHRPP does not define, or chooses not to define, what a "full protocol" or "protocol summary" includes. It is the IRB's responsibility to define what is the "relevant information to determine whether the proposed research fulfills the criteria for approval." In such cases, the study protocol related classes from BRIDG were reused. Since BRIDG focuses mainly on modeling clinical trials (either interventional or observational), it is missing the category of retrospective studies, which use existing health data for secondary analysis. As a result, some entities or attributes in the BRIDG model describing clinical trial protocols are not suitable for describing retrospective studies or social and behavioral science studies. For example, planned study subject experience or subject enrollment is not suitable for secondary analysis studies. Study arms and epochs are not applicable to survey studies. Therefore, when adapting the BRIDG model, the organization of some BRIDG classes was rearranged to properly describe all kinds of human subject research.

Third, with a draft version of the IRB entity-relationship model developed from the first two steps, domain expert reviews were conducted to further refine the model. Not all entities or attributes about study protocol in the BRIDG model are interesting to IRB. Some of them might be important as a blueprint for study conduct or for meeting report requirements from sponsors, but they are not related to the IRB evaluation. Such information elements were eliminated in the IRB model through expert review and only necessary study protocol information was included so that IRB members would not spend

time reviewing information that were not important to their decision making.

Two IRB domain experts were interviewed for the model reviewing. Although domain analysis models such as BRIDG and the IRB DAM were designed to represent a higher abstraction of the domain and be understandable by domain experts, I still found the UML Class Diagram was not intuitive to all domain experts, especially for those without any modeling or informatics background. To make the model more understandable to domain experts, a concept map was derived from the IRB Class Diagram. Concept maps are graphical tools for organizing and representing knowledge. They include concepts, usually enclosed in circles or boxes, and relationships between concepts indicated by a connecting line linking two concepts.[76] Concept maps were originally developed as learning tools in cognitive psychology. However, they are now popularly used in knowledge representation and sharing. Compared to the UML Class Diagram, the concept map includes only the entities in the domain and their relationships without further details such as attributes or data types. The map shows the big picture of the IRB domain and looks much simpler than the UML Class Diagram. Concept maps can use colors to distinguish the central concepts from others, which makes the diagram more readable. One of the interviewed domain experts, with a background in informatics, directly reviewed the UML Class Diagram and the autogenerated documentation, which included a detailed description of each class, attributes, and relationships in the model. The other domain expert does not have an informatics background and was presented with the IRB domain concept map first, and then the UML Class Diagram and the autogenerated documentation.

Both domain experts independently marked each class and associated attributes as

"yes" meaning "to include" or "no" meaning "not to include." After the independent reviews were done by both domain experts, results were consolidated and conflicts were identified. A separate review session was conducted with both domain experts, in which consensus was achieved. The review results were incorporated with both the concept map and the UML Class Diagram.

The final step was to map the information model to real-world IRB application systems to validate its comprehensiveness, which will be discussed in Chapter 5.

4.2.3 Interaction Architecture Modeling

The interaction architecture modeling phase for the IRB DAM differs from the architecture modeling introduced by the FODA method. The FODA architecture modeling focuses on architectural design within the domain application whereas the IRB architecture modeling focuses on architectural design of interactions between the domain application and external systems. This is determined by the different goals of the two domain analysis methods. Since the IRB DAM is mainly designed to achieve system interoperability, the interaction architecture modeling addressed this goal by describing interconnections and information exchange between e-IRB systems and other clinical research information systems.

The interaction model, which is produced from the interaction architecture modeling phase, was represented using the BPMN Collaboration Diagram. BPMN is a standard Business Process Modeling Language (BPML) that is intended to provide a notation readily understandable by all stakeholders. It was developed by consolidating existing notations such as the UML Activity Diagram and Enterprise Distributed Object Computing (EDOC) Business Processes. BPMN is capable of describing the business

processes from different perspectives including the control flow perspective, the organizational perspective and the data perspective.[62,77] The BPMN Collaboration Diagram is one of the three submodels supported by BPMN and is the most suitable for describing the interactions between different systems (participants) using *Pools* and the message exchange between the participants using *Message Flows*. Therefore, BPMN Collaboration Diagram was chosen to model the interactions between e-IRB systems and other clinical research information systems.

The interaction model was developed based on the analysis of current clinical research workflow and identification of potential system integration scenarios. It was not designed to cover all system interactions but illustrated only a few obvious ones for demonstration purposes.

4.3 Modeling Results

This section presents the domain modeling results, including the structural IRB entity-relationship model, the behavioral business process model, and the interaction model. A preliminary effort at building a domain vocabulary is also described.

4.3.1 Concept Map

Although the IRB concept map that was derived from the UML Class Diagram was originally used for facilitating domain expert reviews, it has the advantage of representing domain concepts and relationships in a more straightforward and simple manner. It is an even higher abstraction of the domain and it helps the business analyst to quickly grasp the key elements in a complicated domain from a modeling perspective before examining the details about the attributes and their data types. This was discovered

during the modeling process and can be applied to other modeling efforts in the future. The concept map was developed using the CmapTools software (available for download at: http://cmap.ihmc.us). Each box in the diagram is a concept representing a physical or informational entity in the IRB oversight domain. The links between boxes represent the relationships between concepts. The concept map for the IRB oversight domain is comprised of 97 concepts and 132 relationships in total. As shown in Figure 5, it starts with the most general concept at the top of the diagram "Application," which generalizes three types of applications: IRB applications, ancillary applications, and regulatory applications.

Within the set of IRB applications are four types of applications: 1) Human subject research is the most common type of application reviewed by IRBs and the current IRB DAM focuses only on this type of application; 2) Research projects that do not involve human subjects are not mandated to be reviewed by the IRB according to federal regulations. However, some IRBs still require investigators to submit a nonhuman subject research application and the IRB will make the final decision; 3) Emergency use [78] defined by the FDA as the one-time use of an investigational drug or device for a single participant in a life threatening situation and is exempt from prior IRB review and approval, provided that the emergency use of a test article is reported to the IRB within 5 working days of the date of the emergency use; 4) FDA regulations require local IRB approval before use of a Humanitarian Use Device (HUD), which is a device that is intended to benefit patients by treating or diagnosing a disease that affects fewer than 4,000 individuals in the United States per year.[79] The emergency use of a test article and use of HUD are not considered to be research according to the HHS definition. Since

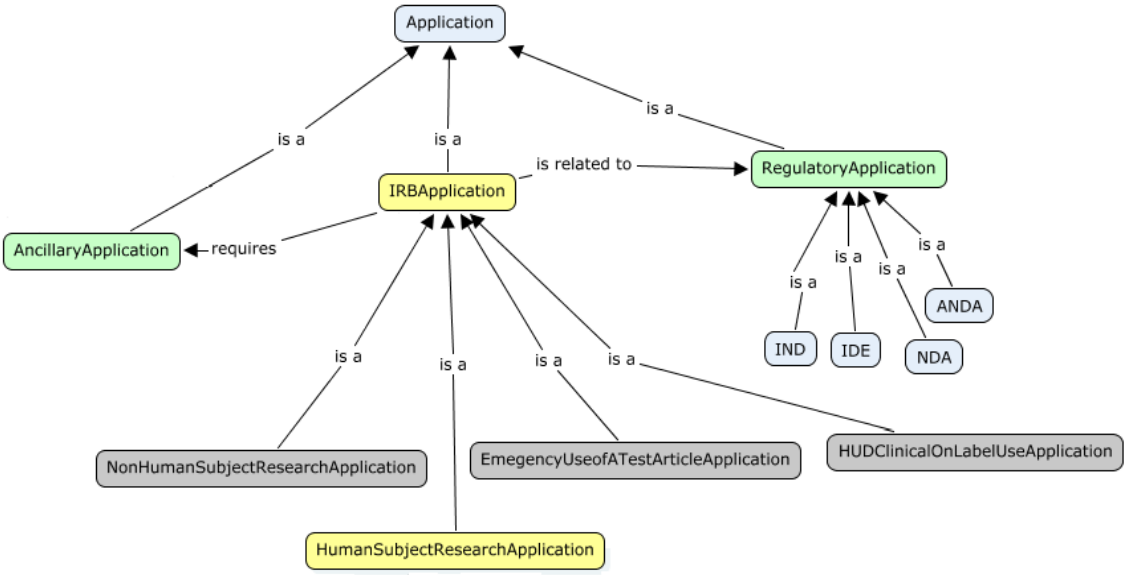

Figure 5. Top-level Concepts Extracted from the IRB Concept Map

our major goal of building the IRB DAM is to streamline the research workflow by promoting interoperability among clinical research information systems, the latter two types of applications are considered as future work. Nonhuman subject research applications are subject to local policy. Therefore, such details were not included in the core model; however, these three types of applications were defined in the concept level in the current IRB model and details can be modeled in the future.

Ancillary applications are submitted to oversight committees other than IRB such as radiation safety review committees, scientific review committees, conflict of interest committees, data and safety monitoring committees, etc. Many IRBs require ancillary application approvals before providing the final IRB approval. These ancillary oversight committees may or may not require extra information besides the standard IRB application. The IRB model is not intended to include all ancillary applications since

many of them are required by local policies. The model, however, the model is designed to be able to support any extension by local IRBs.

Regulatory applications are mandated by regulatory authorities such as the FDA. For studies involving investigational drugs or significant risk devices, regulatory applications such as investigational new drug (IND) or investigational device exemption (IDE) applications are required to be reviewed and approved by the FDA, in addition to the IRB approval. Although it is not the IRB's responsibility to ensure appropriate regulatory approvals are obtained, many IRBs require investigators to submit the IND number or FDA support letter in the IRB application. Such requirements were also considered as local policies and not included in the core model. The model, however, is designed to support these requirements by defining constraining relationships between an IRB application and regulatory applications.

The IRB concept map covers eight core areas (Figure 6), including Study Protocol core, Planned Study Administrative Activity core, Application Amendment and Renewal core, Application Status core, Unanticipated Problem Report core, PHI Authorization core, Ancillary Application Dependency core, and Regulatory Application Dependency core. The key concepts in each core are highlighted in green.

The Study Protocol core represents the informational entities that pertain to the plan of a human subject research. It is essential for evaluation of the study's benefits and risks. Example concepts in the Study Protocol core are planned study site, study subject selection, study conditions, financial sponsors, and planned study activities including any observational and interventional procedures. Many of the concepts in this area were adopted or adapted from the BRIDG model. However, a notable difference is the study

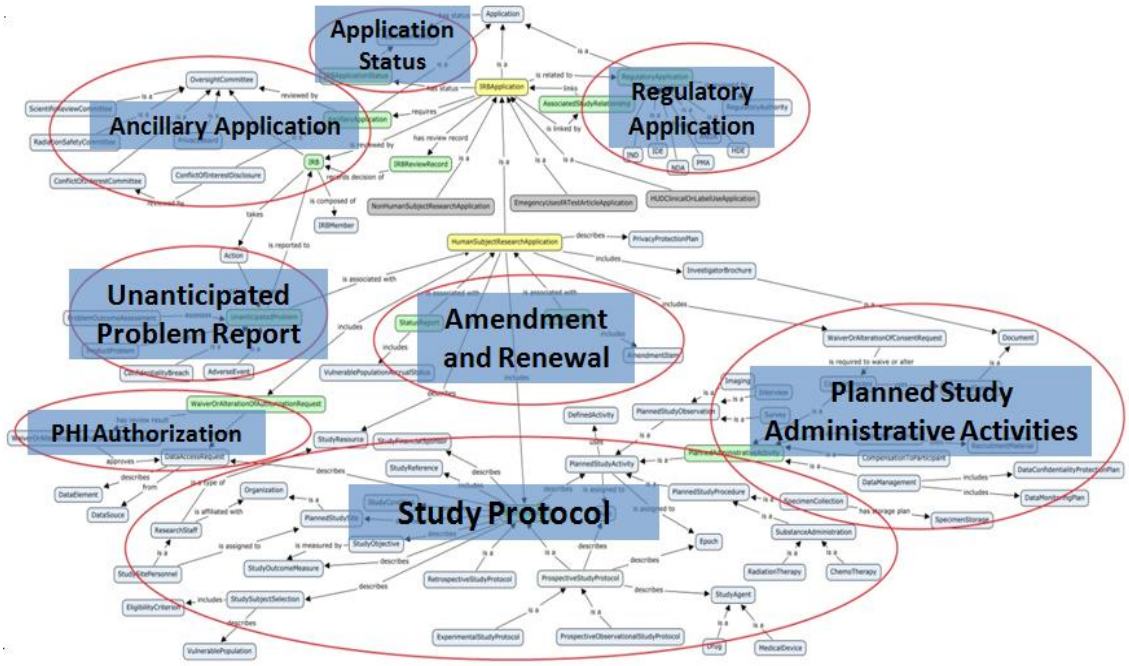

Figure 6. An Overview of the Concept Map for the IRB Oversight Domain

protocol typology specially designed for the IRB oversight domain. A classification of study protocols is needed because different study types need different informational elements to describe the study. For example, a retrospective study uses only existing health data and does not involve any physical contact with study subjects, which does not require screening or interventional procedure plans in the study protocol. In contrast, a prospective study protocol should describe in detail the recruitment process and all observational or interventional procedures that will be applied to study participants. A well-designed study protocol typology, with relevant information elements defined for each study type, can facilitate "smart" form design in e-IRB systems so that investigators do not need to answer inapplicable questions. It should also address the key points such as the level of risk considered by the IRB during the review process.

There are a number of existing classification schemas that have been defined in the

clinical research domain to suit different purposes. For example, a common classification categorizes clinical research as experimental, quasi-experimental or nonexperimental (also referred to as observational research). Carini et al. developed the Study Design Typology as part of the Human Studies Database project[80] to standardize the classification of study designs in human research. It categorizes human studies into quantitative studies and qualitative studies with a focus on further classification of quantitative studies. This typology is designed to support the critical appraisal of evidence of human research and classification of new and ongoing research for scientific portfolio management and analysis. It is designed to distinguish only interpretive concerns among studies. There is no single classification schema that meets the needs of all scenarios. The existing study classification schemas are not best suitable for IRB review purposes. The IRBs' goal is to protect human subjects in research, and the review process focuses on the evaluation of the risks to participants. Knowing if the study is quantitative or qualitative does not immediately help IRB understand the risks of the study since quantitative studies may be retrospective and only cause informational risks to study participants whereas qualitative studies may involve interviews with vulnerable populations that raise ethical concerns. Therefore, based on literature reviews and consulting with IRB domain experts, the IRB model developed a study protocol classification schema specifically for the IRB domain (Figure 7). Study protocols are categorized into retrospective study protocols and prospective study protocols depending on the time perspective of the study design. Prospective study protocols can be further categorized into experimental study protocols and prospective observational study protocols. This categorization gives IRB a straightforward indication of the risk level

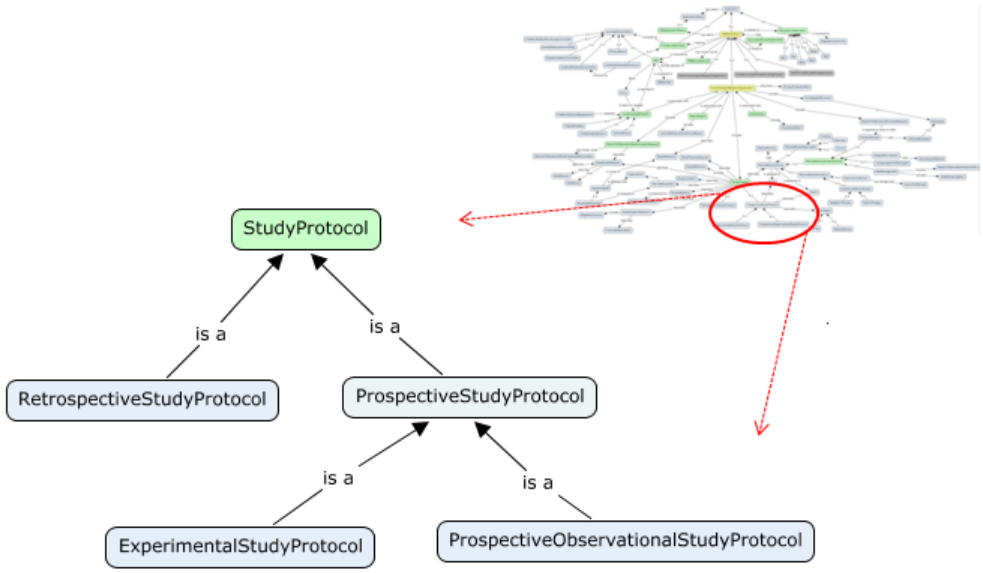

Figure 7. The Study Protocol Typology for the IRB Oversight Domain

associated with the study. Further detailed study design information for experimental or observational studies are specified by a coded attribute "design configuration," the value set of which can be derived from Carini's Study Design Typology. The informational elements describing each study protocol type including those from the Protocol Representation subdomain of the BRIDG model were rearranged so that the most general study protocol type only includes informational elements that are applicable to all types of studies and more specialized informational elements were defined for more specific types of study protocols.

The Planned Administrative Study Activity core represents those activities that are not directly related to the analysis of study outcome such as participant recruitment procedures, compensation to study participants, and informed consent processes, etc. Strictly speaking, these activities are part of the planned study activities described in the study protocol. However, since these administrative activities can raise significant ethical

concerns such as equitable subject selection and respect for person, IRBs always require investigators to specify the details of these activities, which may not be included in a typical study protocol, which focuses more on study outcome related activities.

The PHI Authorization core covers the informational entities that are related to the IRB (or Privacy Board) as regulated by the HIPAA Privacy Rule. For example, the request for waiver or alteration of authorization describes the PHI data elements to access, as well as the justification of such access. Documentation of the IRB's review and approval of the request is also required by the HIPAA Privacy Rule. In fact, what kind of PHI access has been approved by IRB is essential to achieve automated access control in SUHD systems and effective IRB oversight of PHI disclosure. The authorization document itself was not included in the core model since it does not require review and approval by the IRB, according to the HIPAA Privacy Rule.

The Unanticipated Problem Report core defines the informational elements that should be reported to the IRB and corresponding action taken by the IRB. The unanticipated problem report is not only useful for protecting study subjects from avoidable harm during the study, but also helpful for the IRB to reevaluate the study during the continuing review process. This portion of the model adopted some concepts from the Adverse Event subdomain of the BRIDG model. Since adverse events are often required to be reported to other bodies such as a local Data and Safety Monitoring Committee, the sponsor or the FDA, a standardized format can facilitate automated report and avoid duplicate report preparation efforts. However, not all adverse events are reportable to the IRB. On the other hand, there are other types of unanticipated problems besides adverse events, such as a data breach or product problem.

The Application Amendment and Renewal core defines information entities such as amendment items and status report that are related to the IRB's continuing review. A standardized status report can be automatically generated from existing information stored in a CTMS by defining certain report parameters.

The Application Status core represents information related to the status of an IRB application and can be shared with other clinical research information systems to automate a streamlined workflow. This can eliminate the needs for manually delivering paper-based IRB approval letters to different stakeholders in the research domain.

The Ancillary Application core and the Regulatory Application core were already described as peer concepts of the IRB application in previous sections.

4.3.2 UML Class Diagram

The UML Class Diagram is the detailed entity-relationship model that describes not only the concepts and relationships, but also attributes and their data types in the IRB oversight domain. Each box in the Class Diagram is a class that corresponds to a concept with the same name in the concept map diagram. There are attributes listed in each box describing the properties of a concept. Each attribute is bound to a data type defined in the HL7 Version 3 Data Type Abstract Specification (Release 2).[81] Each data type defines the structural format of the data and influences the set of allowable values an attribute may assume. HL7 abstract data type was chosen because it was specifically developed for the healthcare domain and is independent from representational and operational concerns or specific implementation technologies, which makes it suitable for a domain analysis model. The complete model in UML diagrams can be accessed online at http://irb-dam.bmi.utah.edu.

4.3.2.1 Legend

Unlike the concept map diagram, which is color coded by the coordination of the concept in the IRB oversight domain, the UML class diagram is color-coded by the information sources that led to the creation of the class, as depicted in Figure 8. Classes that were adapted or adopted from the BRIDG model use their original color code and a bold red border was used to distinguish them from other classes. This will make the harmonization process with the BRIDG model easier in the future. Some classes that were adapted from the BRIDG model were also recommended in regulatory guidelines, but were color coded as BRIDG classes for the same reasons. The attributes of the AAHRPP recommended classes were either from the AAHRPP guidelines directly when available such as StatusReport or defined by referring to related OHRP or FDA guidelines or CFRs. The Miscellaneous class category includes classes that were created based on modeling best practices (e.g., parent and child classes in a generalization relationship) or based on domain expert review feedback. Table 2 provides a summary of the number of classes from each knowledge source. The knowledge sources were recorded in the "Note" field of the class (Figure 9). This information is used for future reference. For example, when the regulations or guidelines change, impacted classes can be easily identified and then updated. Most class and attribute names are self-explanatory and a brief definition for each class and attribute was documented in the "Note" field.

4.3.2.2 A Glance at the IRB UML Class Diagram

This section describes a few representative classes in the UML Class Diagram to explain the modeling techniques and demonstrate the modeling result. The complete model can be accessed online at http://irb-dam.bmi.utah.edu.

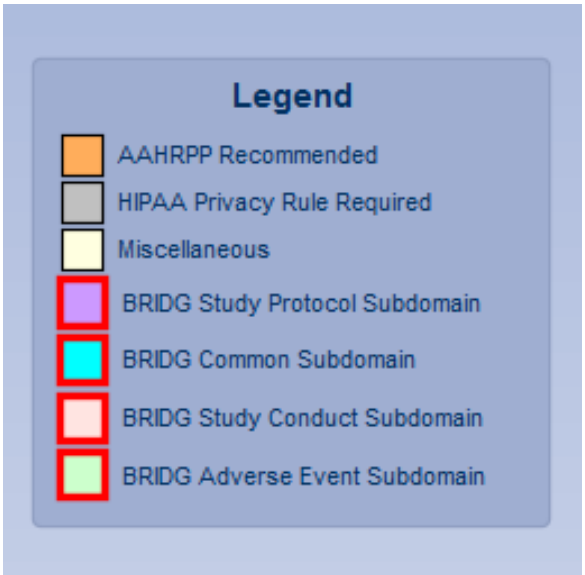

Figure 8. Legend for the UML Class Diagram

Table 2. Summarization of the Number of Classes by Knowledge Source

Source	Number of Classes	Percentage
Regulatory Guidelines (Mainly AAHRPP)	30	30.9%
HIPAA Privacy Rule	4	4.1%
BRIDG Model	21	21.6%
Miscellaneous	42	43.3%
Total	97	100%

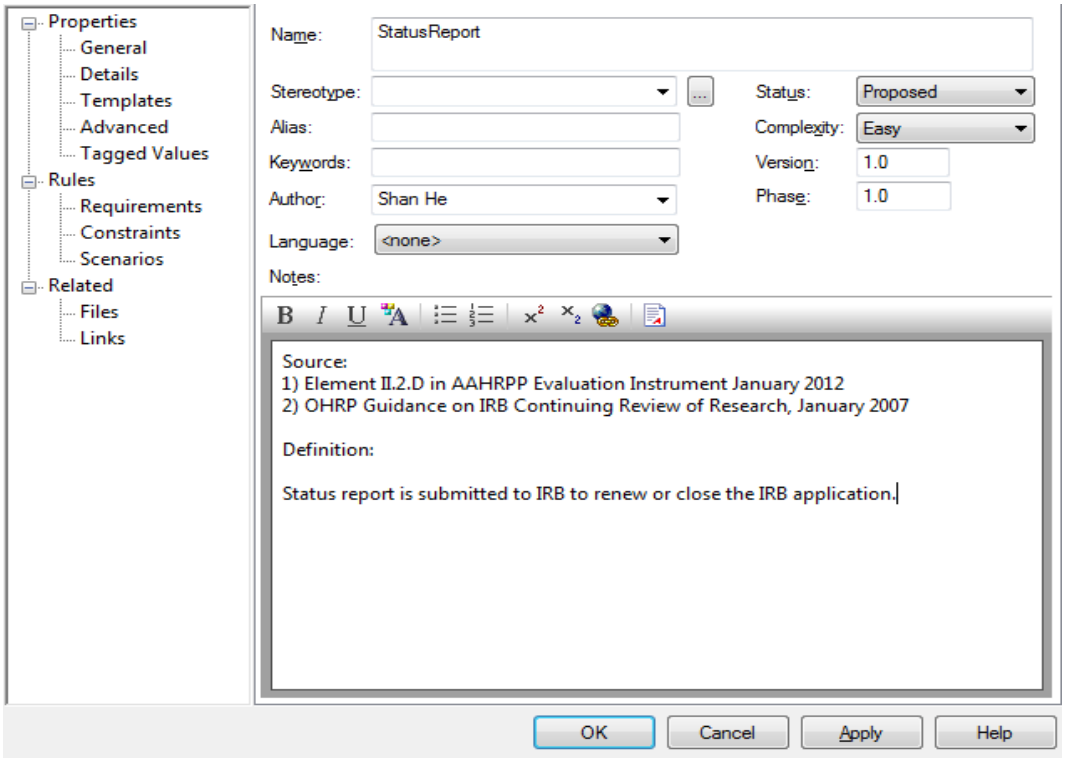

Figure 9. The Note Field of a Class Documenting the Knowledge Source

Figure 10 provides a close look at part of the Class Diagram as an example. Highlighted with red circles, the diagram shows the class name on the top of each box. Each class has certain attributes listed under the class name. There are two parts composing an attribute: attribute name (e.g., *applicationType*) and attribute data type (e.g., *CD*). CD is the symbol for *ConceptDescriptor* defined in HL7 abstract data type specification. It refers to a concept defined in a value set. Compared to the *string* (ST) data type, which represents free text values, the CD data type allows only coded values that are machine-interpretable. There are arrows linking boxes, which represent the relationships between classes. For example, the generalization relationship (which sometimes is referred to as *is-a* relationship) is a relationship in which one class (the child) is based on or a specialization of another class (the parent). The UML graphical

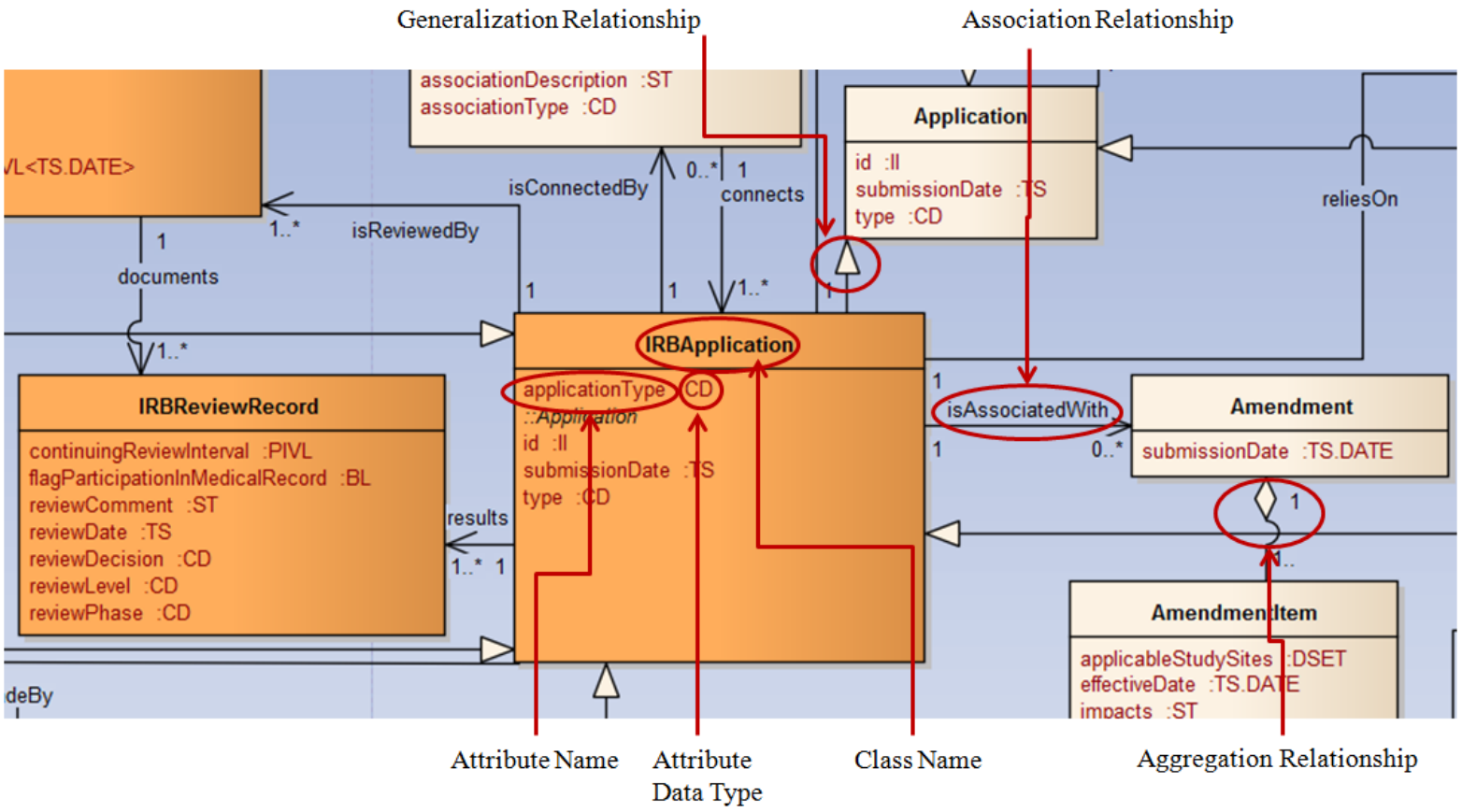

Figure 10. Illustration of the UML Class Diagram in Detail—Modeling Constructs

representation of a generalization relationship is a hollow triangle shape on the parent class end of the line that connects it to one or more child classes. The child class *IRBApplication* inherits all the attributes defined in the parent class *Application*. Another type of relationship is the association relationship, which describes any type of association between two classes. In this example, *Application* is associated with *Amendment*. There is a third type of relationship called aggregation relationship, which shows a class as a part of another class. The UML graphical representation of an aggregation relationship is a hollow diamond shape on the whole class end of the line that connects it to one or more part classes. In the example, an aggregation relationship connects the whole class *Amendment* and the part class *AmendmentItem* meaning *AmendmentItem* is part of *Amendment*.

A study protocol may change during the administration of the study. The BRIDG model defined the class *StudyProtocolVersion* to represent the content of a study protocol at a particular point in time. However, in the IRB oversight domain, not only the changes of the study protocol, but also why the changes were made and how these changes may impact the study participants, are important for IRBs in re-evaluating the study. Therefore, metadata about changes such as applicable study sites, impact, and reasons for change should be defined for each amendment item. An amendment item can be any previous submitted or newly added information with regard to the study protocol, informed consent and any other supporting material. The view of the study protocol at a particular point in time is considered to be an implementation level issue, which can be generated automatically based on merging the original study protocol and amendments.

Some of the attributes in the *StudyProtocol* class were borrowed from the BRIDG

StudyProtocolVersion class, with IRB-irrelevant attributes removed and clinical trial only properties reattributed to children classes. For example, "accrualReportingMethodCode," which is a coded value specifying the technique that is used for reporting study subject accrual data to the study sponsor is not relevant for IRB review. Therefore, it is not included in the IRB model. The "plannedStudySubjectExperience," which is a description of what the study subject can expect to experience over the course of the study, applies only to prospective studies. Therefore, this attribute was moved from the *StudyProtocol* class to the *ProspectiveStudyProtocol* class. Some new classes and attributes related to study protocol that were not included in the BRIDG model but are important for IRB review were added to the IRB model (e.g., *DataAccessRequest*, *VulnerablePopulation, etc.*). A separate class *StudySubjectsSelection* was created to describe all study subject related attributes including inclusion criteria, exclusion criteria, target number, and supplementary details about vulnerable population (Figure 11). Even though the "inclusionCriteria" and "exclusionCriteria" attributes are defined as strings that are designed for human review, their values could be automatically derived from the structured *EligibilityCriterion* class especially for medical criteria.

Structured or computable representation of eligibility criteria plays an important role in facilitating automated research participant screening, clinical evidence application, and clinical research knowledge management. There are a number of active research projects on this topic.[82–85] A deep analysis of computable eligibility criteria is beyond the scope of this dissertation. A simplified eligibility criteria expression model that combines clinical statements with logical connectors (e.g., and, or, not) was constructed in the IRB model with an intention to represent most of the study eligibility criteria without building

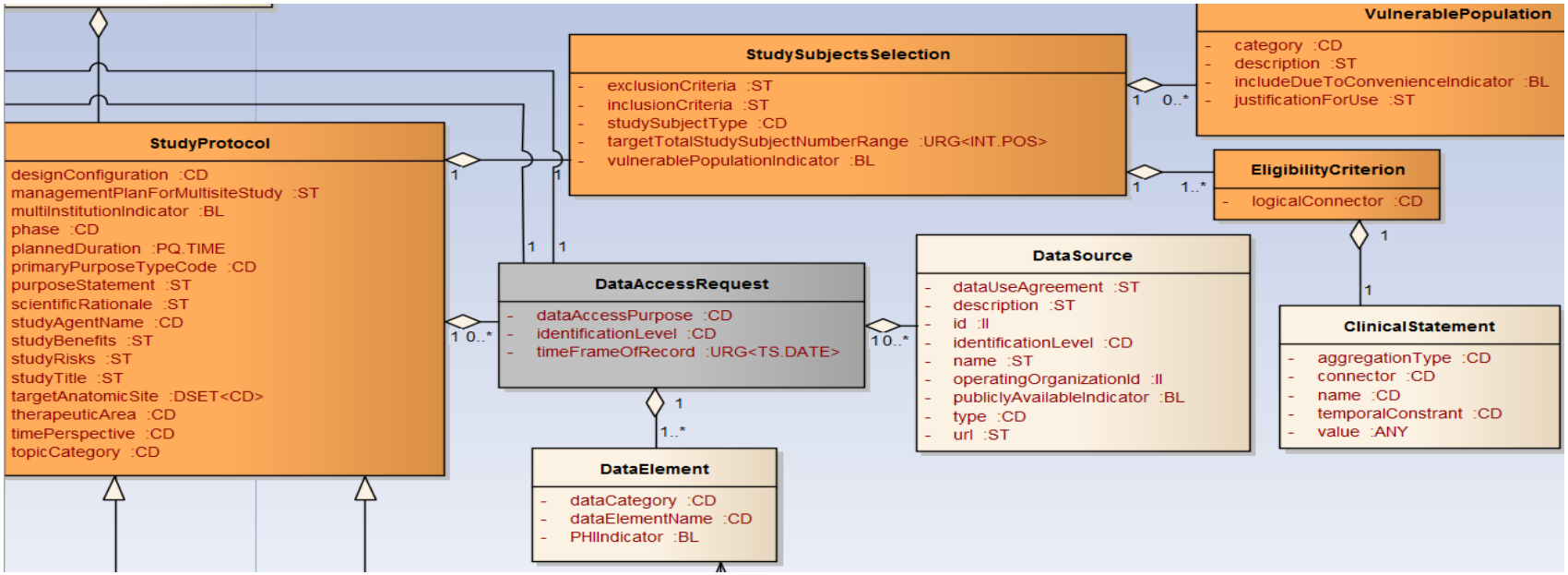

Figure 11. Illustration of the UML Class Diagram in Detail—Study Subject Selection and Data Access Request

an overly sophisticated model. Any eligibility criterion that could not be expressed in the model, especially those nonmedical criteria such as "access to telephone," can be stated in the free-text eligibility criteria attributes.

There are also cases where the class definition in BRIDG is not quite clear or has a different perspective from the IRB's interest. In these cases, the BRIDG class was renamed and redefined for the IRB domain. For example, BRIDG defined the class *StudyLegalSponsor* while IRB is specifically interested in the financial sponsor of a study. Although sometimes these two types of entities overlap with each other, the IRB model clarified this case by defining the class *StudyFinancialSponsor*.

The BRIDG model used a naming convention for attributes with data type CD—all such attributes have a suffix "Code" (e.g., designConfigurationCode, phaseCode). However, although the HL7 data type CD represents any kind of concept that is usually defined in a code system, it is an abstract data type that is composed of multiple properties such as code, codeSystem, codeSystemName, codeSystemVersion, displayName, originalText, translation, and qualifier. Moreover, in cases of an exception value, the CD need not contain a code but only the original text describing that concept. Attributes named with a "Code" suffix may be confusing to domain experts since they might think it is simply a computer-interpretable code. Therefore, in the IRB domain model, the "Code" suffix was removed from attributes borrowed from the BRIDG model.

As shown in Figure 11, the IRB model defines the data access request related classes for studies that require access to existing data for secondary analysis or participant recruitment. Neither the BRIDG model nor the AAHRPP guidelines cover this category of information. The DataAccessRequest class is associated with other classes including

DataElement and DataSource. DataElement describes the data categories to be requested (e.g., demographics, conditions, procedures, drug exposures, etc.) and specific data element names when necessary (e.g., gender, race, year of birth, month of birth, day of birth for the demographics category). DataSource describes where the requested data are from. This can be implemented in an e-IRB system by providing a list of data sources in an institution that are under the jurisdiction of the IORG. The attributes defined by the DataAccessRequest class are listed in Table 3.

4.3.3 Concept Domain—Terminology Binding

Our major goal in developing an IRB DAM is to achieve interoperability between e-IRB systems and other CRI applications. The IEEE Glossary defines interoperability as "the ability of two or more systems or components to exchange information and to use the information that has been exchanged."[86] Implementation of the IRB entity-relationship model (information model) enables syntactic interoperability by specifying the structure of information being exchanged between different systems. However, to make the meaning of the content being exchanged understandable to the receiving system (semantic interoperability), an information model needs to be bound to value sets where each value set consists of one or more of the coded concepts. In HL7, the term "Concept Domain" is used to refer to "a named category of like concepts that will be bound to one or more coded elements."[87] Every attribute with the HL7 data type CD in the IRB model should be bound to a Concept Domain, which is in turn bound to one or more value sets. Such value sets can be defined from scratch or by adopting existing terminologies or code systems if available. I defined the value sets for a few attributes in the model as a preliminary IRB domain vocabulary specification effort.

Table 3. Attributes Defined for the *DataAccessRequest* Class

Attribute Name	**Definition**	**Example Values**
dataAccessPurposeCode	the purpose of the data access request	potential participant identification; secondary analysis
identificationLevelCode	the identification level of the data requested (this can be derived from *DataElement* values)	de-identified; limited data set; identifiable;
rangeOfNumberOfRecords	the range of the number of records requested	500-600
timeFrameOfRecords	the time frame of the records requested (can be derived from *EligibilityCriterion*)	1990-2010

Figure 12 shows an example value set defined for IRB application status and Figure 13 shows an example value set defined for the vulnerable population category. However, a formal and comprehensive IRB domain vocabulary specification is considered as future work since it needs a significant amount of iterative development and evaluation effort.

4.3.3.1 Research-Oriented Health Data Representation Model

This dissertation is specifically focused on defining the values sets for data access request related attributes such as *dataCategory* and *dataElementName* because a prototype implementation of the IRB model as part of the dissertation involves transmitting this information between a SUHD system and an e-IRB system. In the last decade, various developments have occurred to specify clinical models (sometimes called templates or archetypes) to represent clinical data with structured data elements in the healthcare domain to address a multitude of purposes.[88] These models are developed mainly to address the storage and exchange of electronic patient records in healthcare settings. They are not suitable for representing data elements for secondary use of health data in the clinical research domain. In the IRB model, I adapted the Common Data Model (CDM) Version 4 from the Observational Medical Outcomes Partnership (OMOP).[89] The OMOP CDM is designed to support research by standardizing the structure and content of various health data sources. Although it was originally developed to support drug safety surveillance research, it has evolved to support clinical research in general. The entity-relationship diagram (ERD) of the OMOP CDM is represented in implementation levels for relational database table design. The OMOP ERD was translated to a high level domain information model (logical data model) represented as UML Class Diagram with implementation-related attributes removed. A few classes were

.

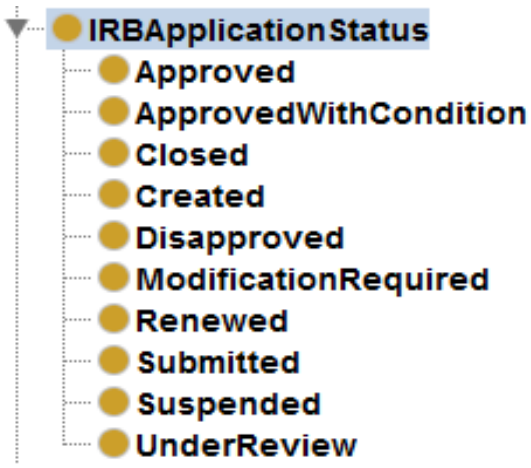

Figure 12. Example Value Set for *IRBApplicationStatus*

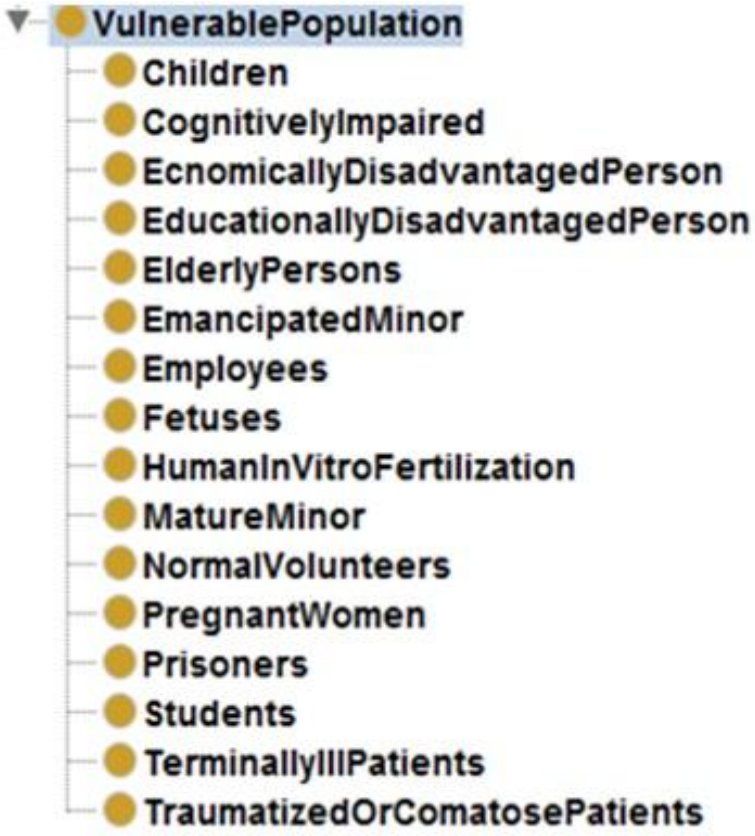

Figure 13. Example Value Set for *VulnerablePopulation*

added to the translated OMOP model based on domain expert consultation and this formed our Research-Oriented Health Data Representation (ROHDR) model. For example, the *Observation* class was extended to represent specific types of observation such as *LabObservation*, *RadiologyObservation*, *PathologyObservation*, *PhysicianNote*, *Allergy Observation*, *Vitals*, etc. A detailed documentation of the ROHDR can be accessed online at http://irb-dam.bmi.utah.edu. The *DataAccessRequest.dataCategory* attribute is bound to a value set consisting of all class names in the ROHDR model. The *DataAccessReques.dataElementName* attribute is bound to a value sets including the attributes defined for each data category

4.3.4 The Business Process Model

The behavioral aspect of the IRB DAM is the business process model represented using the BPMN Process Diagram and the UML State Machine Diagram. The current business process model focuses only on the high-level IRB review processes. Details about the IRB review workflow design are considered as implementation level issues. Figure 14 illustrates the initial IRB application and review process. Figure 15 illustrates the IRB continuing review process. Three roles (participants) are involved in both processes: investigator, e-IRB system, and the IRB reviewers.

Besides the two business process diagrams, a state machine diagram was created to represent the status transition of an IRB application as shown in Figure 16. The “In Review” state is a composite state, which is composed of multiple substates as shown in Figure 17.

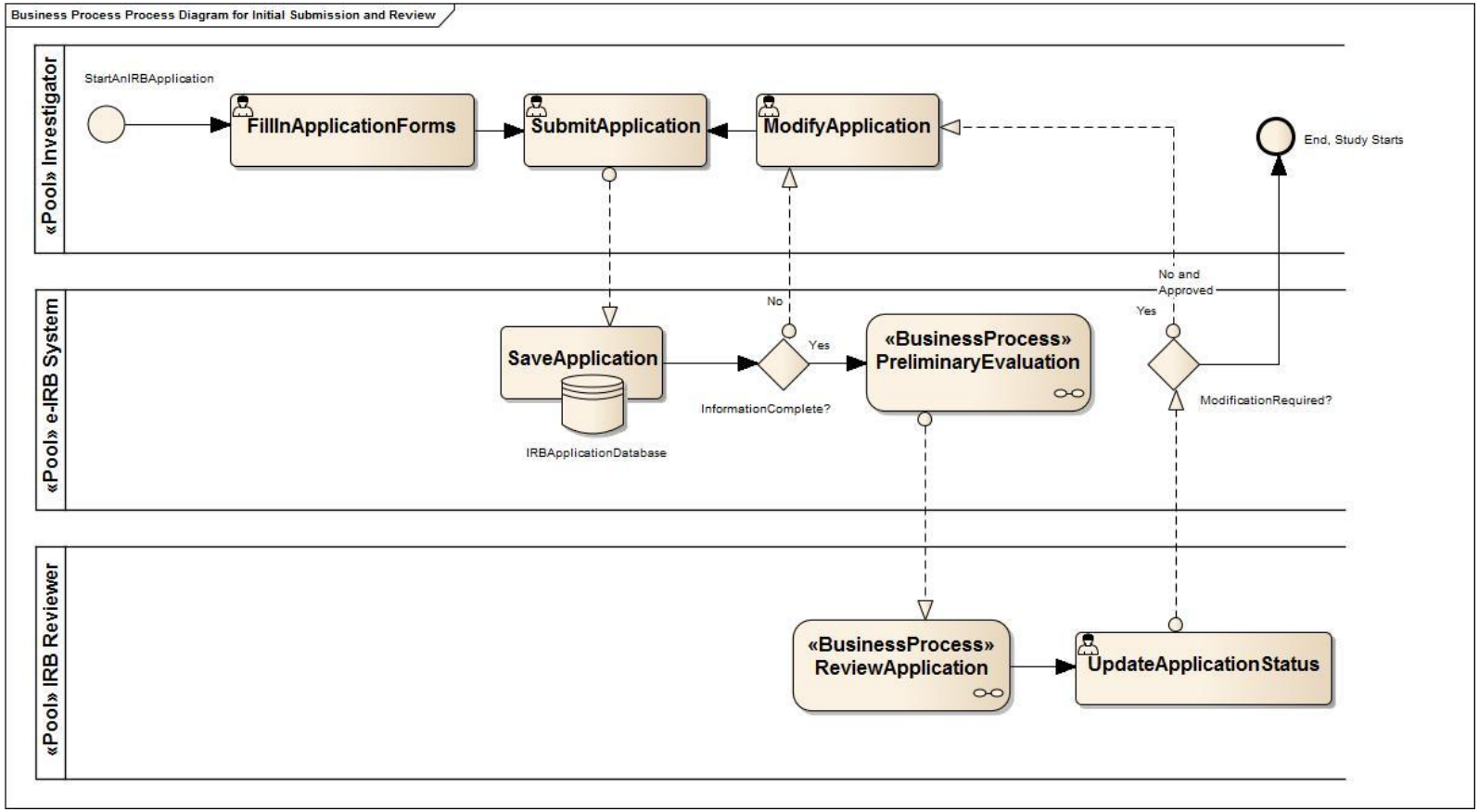

Figure 14. Business Process for Initial IRB Application and Review

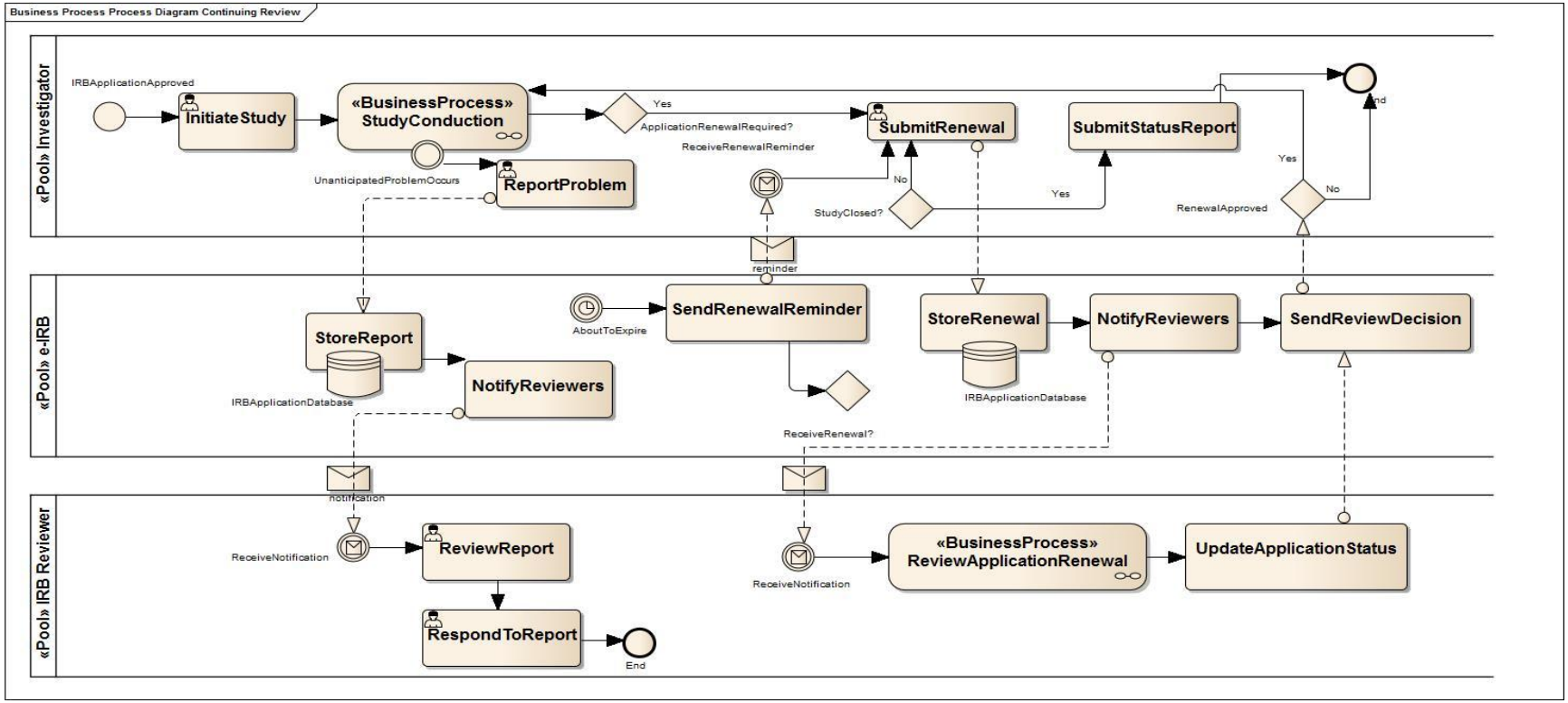

Figure 15. Business Process for IRB Continuing Review

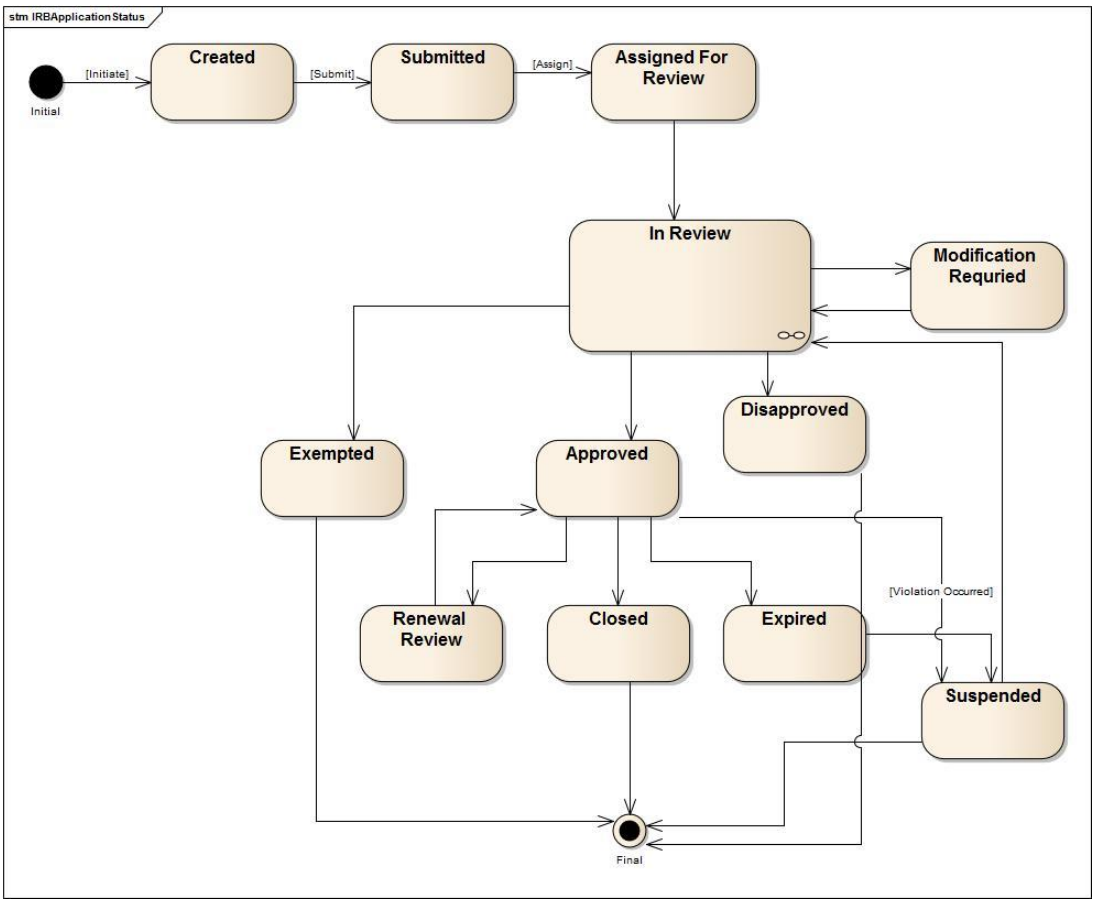

Figure 16. State Machine Diagram for IRB Application Status

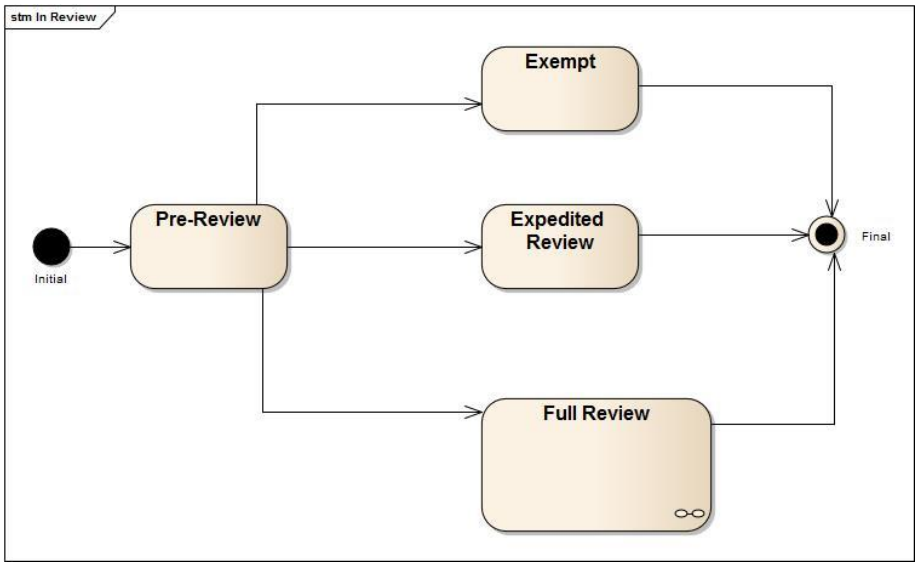

Figure 17. Substate Machine for the Composite State “In Review”

4.3.5 The Interaction Architecture Model

The interaction architecture model is represented using a BPMN Collaboration Diagram. Figure 18 illustrates the interaction model for automated access control on secondary use of PHI as discussed in Chapter 2. Three roles (participants) are involved in the interaction: Investigator, an e-IRB system, and a SUHD system. Similarly, the interaction model for automated protocol extraction from a CTMS to an e-IRB system can be accessed online at http://irb-dam.bmi.utah.edu. Interaction models for integrating e-IRB systems with other CRI systems will be developed in the future as the adoption of the IRB model evolves.

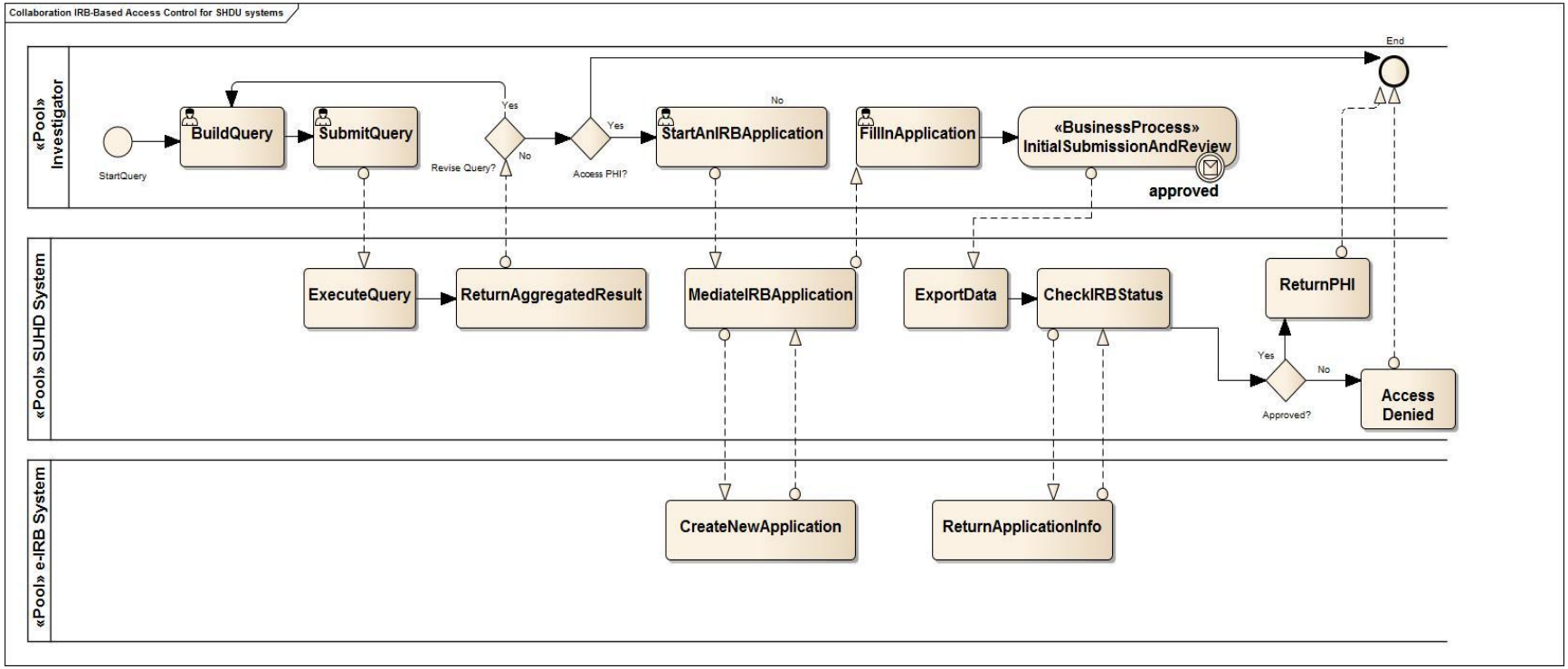

Figure 18. An Example Interaction Model for Automated Access Control on Secondary Use of PHI

CHAPTER 5

MODEL EVALUATION

The structural IRB entity-relationship model was evaluated to validate its support for representing informational elements in an IRB application required at different institutions. The evaluation included comparisons with real-world IRB application systems from five representative institutions. These IRB application systems were chosen because they come from institutions across the nation and each of them is representative of a typical submission method or review model. Table 4 summarizes the five IRB application systems used in the evaluation phase.

5.1 Evaluation Methods

The evaluation was performed from January to March 2013. The most up-to-date Web forms or Word/PDF application templates during that period of time at each institution were used for the evaluation. Most institutions make their IRB application forms in Word/PDF format publicly accessible. For those institutions that need an account to access the application forms, I contacted the IRB office and a temporary access account was created for this evaluation project. For certain e-IRB systems that require institutional accounts, either an account was registered or equivalent Word/PDF application forms were used for the evaluation. Each field defined in the Web form or

Table 4. Summarization of the Five IRB Application Systems

IRB System Name	**Institution**	**System Type /Review Model**
ERICA	University of Utah	Customized commercial e-IRB /Institutional
DISCOVR-E	Vanderbilt University	In-house developed e-IRB/Institutional
Harvard IRB	Harvard Medical School	Word templates submitted via e-mail or hard copy (e-IRB went live on March 18, 2013)/Institutional
$WIRB_{NET}$	Western IRB	Ad hoc e-IRB/Independent Commercial
Central IRB (CIRB)	National Cancer Institute	Mixed submission method combining e-mail and SaaS e-IRB(IRBManager)/Central

Word/PDF templates was extracted as an information item and mapped to the IRB entity-relationship model (referred to as "Model") with one of the eight mapping types as the mapping results (Table 5).

5.2 Evaluation Results

Table 6 summarizes the mapping results for the five IRB application systems. The detailed mapping results for each IRB are summarized and displayed using pie charts and are described in the sections below. The data were analyzed using IBM SPSS Statistic 20.

5.2.1 University of Utah IRB

The ERICA system used by the University of Utah IRB is an institutional e-IRB system and is accessible using a university issued ID. The fields from all Web forms including forms for the initial application, application amendment, application renewals, and unanticipated problem reports were mapped to the Model. There were a total of 280 fields extracted from ERICA. In order to illustrate the mapping results more clearly, five categories of mapping--Exact Mapping, Equivalent Mapping, Partial Mapping, Supportable Mapping, and Derivable Mapping--are considered as certain forms of mapping and combined into a general category (Figure 19). The detailed distribution of each mapping category is illustrated in a separate chart as shown in Figure 20. The mapping results for other IRBs are illustrated in the same way.

5.2.2 Vanderbilt University IRB

The DISCOVR-E system used by the Vanderbilt IRB requires a Vanderbilt University account to gain access to the system. I contacted the Vanderbilt IRB office and obtained a

Table 5. Description of the Mapping Types

Mapping Type	**Definition**	**Example**
Exact Mapping	The form field can be exactly mapped to an attribute of a class in the Model.	"Title of Study" →StudyProtocol.studyTitle
Equivalent Mapping	The form field can be mapped to the Model by combining more than one attributes from one or more classes.	"Describe all the procedures chronologically, from screening/enrollment through study closeout, which will be completed in the research project." →PlannedStudyActivity and all its attributes
Partial Mapping	The Model has a general attribute covering more than one related form fields but lacks the specificity defined in the form fields.	"Request for waiver or alteration of informed consent 1) Purpose of the Waiver Request; 2) Explain why the research could not practicably be conducted without the waiver or alteration; 3) Explain why the research and privacy risk of the research are no more than minimal; 4) Describe the measures you will take to ensure the waiver or alteration will not adversely affect the rights and welfare of the subjects"→ WaiverOrAlterationOfConsentRequest justificationForRequest

Table 5 Continued

Mapping Type	Definition	Example
Derivable Mapping	The form field cannot be directly mapped to a class or attribute in the Model but it can be derived from other attribute(s).	For status report, some IRBs ask “number of participants enrolled since last review” and “number of participants enrolled since the beginning of the study.” Only the first one is defined in the Model and the second can be derived by adding up the previous numbers.
Supportable Mapping	The form field is supported by defining value set(s) for a certain attribute in the Model.	Some IRB asks for study personnel’s information such as contact person, internal staff, external sub-investigator, guest, etc. This type of information is supported by defining the value set for ResearchStaff.role by local IRBs.
Out of scope	The form field is defined according to local regulations or policies and it is intentionally excluded from the core	Studies that involve only de-identified data or limited data set access are not required to be reviewed by IRBs by federal regulations. However, many IRBs require investigators to submit an application

Table 5 Continued

Mapping Type	Definition	Example
	model. However, it is possible to extend the Model to support such local policies.	for such studies and the IRB will determine if the study qualifies for nonhuman subject research
Not Defined	The Model does not have a corresponding class or attribute defined for the form field.	Details about placebo-controlled studies, HIV antibody testing related details.
Unclear	The definition of the form field is not clear.	"Number ineligible for study" about participant information in the continuing review form should be clarified, for example "Number of subjects ineligible after the screen procedure." Some IRB asks very general questions like "How the rights and welfare of participants will be protected?"

Table 6. Summarization of the Mapping Results

Institution	Total num of fields	Exact Mapping	Equivalent Mapping	Partial Mapping	Supportable Mapping	Derivable Mapping	Not Defined	Out of Scope	Unclear
University of Utah IRB	280	23.9%	14.3%	17.1%	21.1%	5.4%	9.6%	7.5%	1.1%
Vanderbilt University IRB	241	14.9%	20.7%	35.7%	7.5%	6.6%	8.7%	4.9%	0.8%
Harvard University Faculty of Medicine IRB	263	20.5%	23.6%	16.3%	10.6%	9.1%	13.7%	2.7%	3.4%
Western IRB	302	13.6%	8.3%	32.5%	18.2%	2.6%	9.6%	14.9%	0.3%
NCI Central IRB	141	17.7%	6.4%	21.3%	22.7%	2.8%	12.1%	16.3%	0.7%

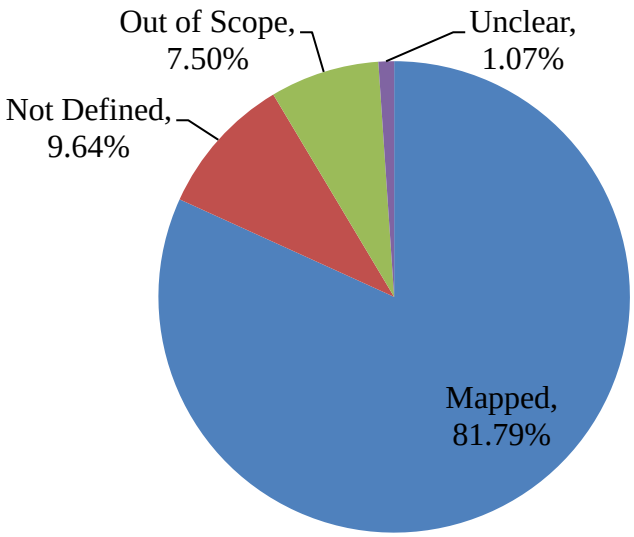

Figure 19. Mapping Results for University of Utah IRB

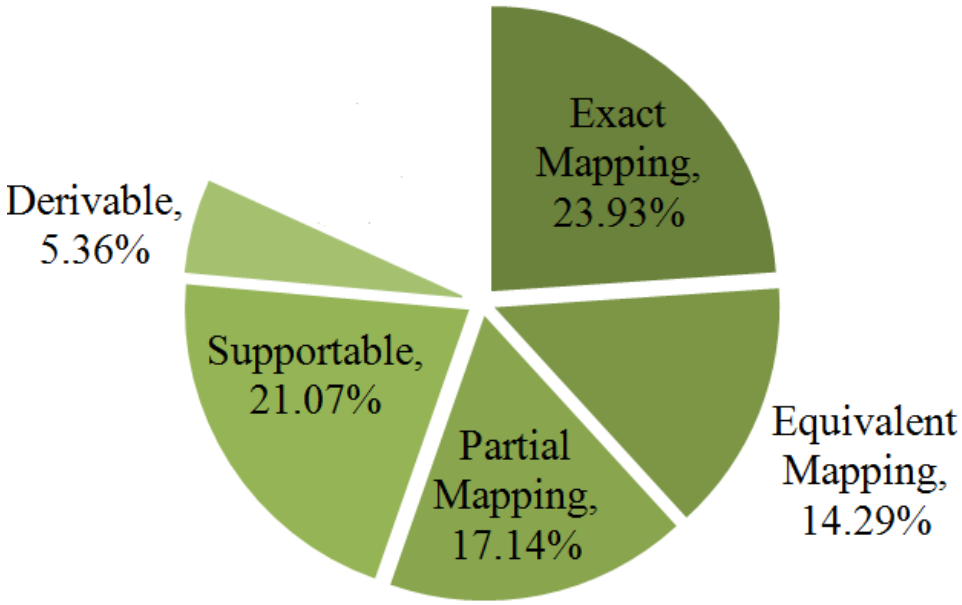

Figure 20. Mapping Details for University of Utah IRB

confirmation that their application forms in Word format contain the same content as the Web forms in DISCOVR-E and are accessible to the public. The 18 total forms for mapping included initial and continuing review application forms for social and behavioral sciences and for health sciences and additional application forms for request for exemption and studies involving medical devices, specimens and data repositories, or vulnerable populations. Duplicate information items in different forms that can be eliminated using "smart" forms in an electronic system were included only once for the purpose of mapping. A total of 241 fields were extracted from the forms. The mapping results for Vanderbilt IRB are illustrated in Figure 21 and 22.

5.2.3 Harvard University Faculty of Medicine IRB

Twelve forms are included from the Committee on Human Studies (the IRB equivalent oversight body) at the Harvard University Faculty of Medicine. The term IRB is used for consistency throughout the dissertation. The following forms were included for mapping: the general form for initial application; extra forms for use of specimen or data; ionizing and nonionizing exposure; use of drugs, biologics, and devices; research involving minors; request for consent waiver form; exemption determination form; unanticipated problem report form; study renewal form; and study closure form. Duplicate fields across forms were included only once. The Harvard IRB switched from Word-based application forms to an e-IRB system in March, 2013.

Because this dissertation research was already underway, it relies on the Word-based system for mapping analysis. The mapping results are shown in Figure 23 and Figure 24.

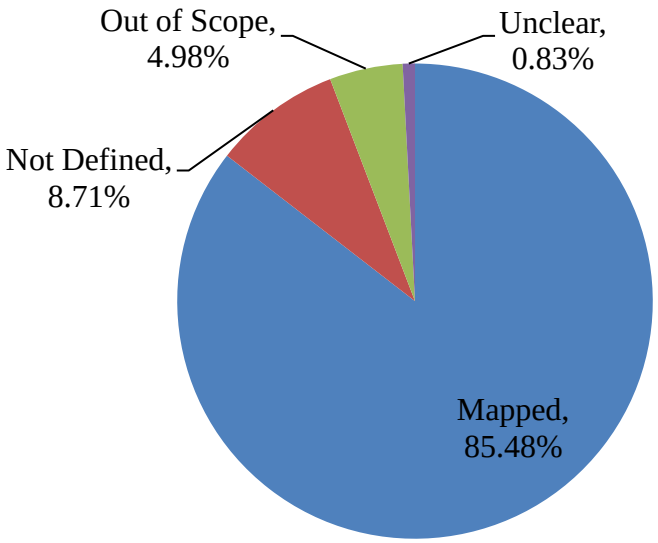

Figure 21. Mapping Results for Vanderbilt University IRB

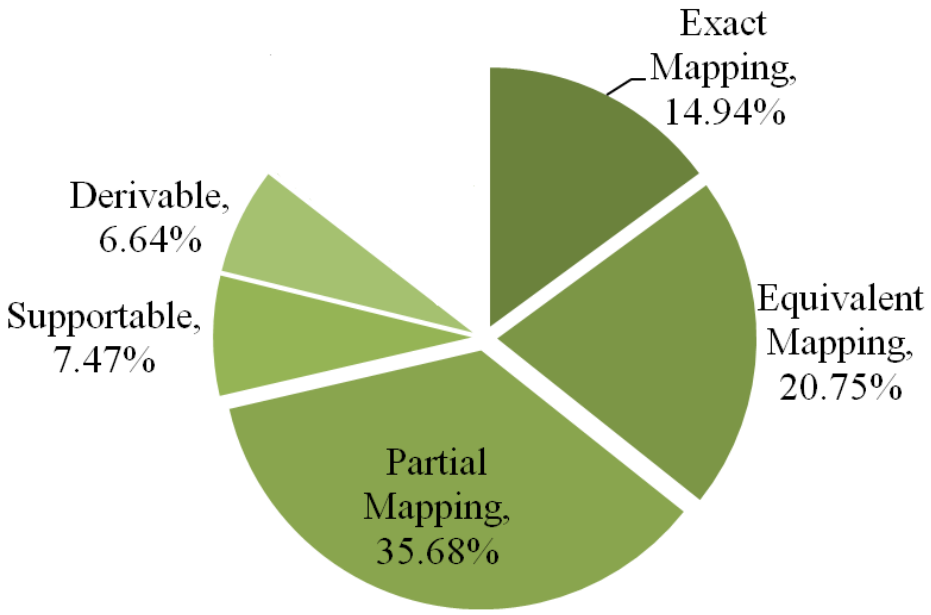

Figure 22. Mapping Details for Vanderbilt University IRB

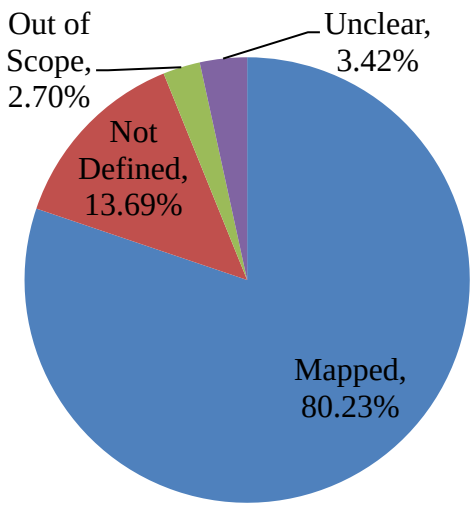

Figure 23. Mapping Results for Harvard Medical School IRB

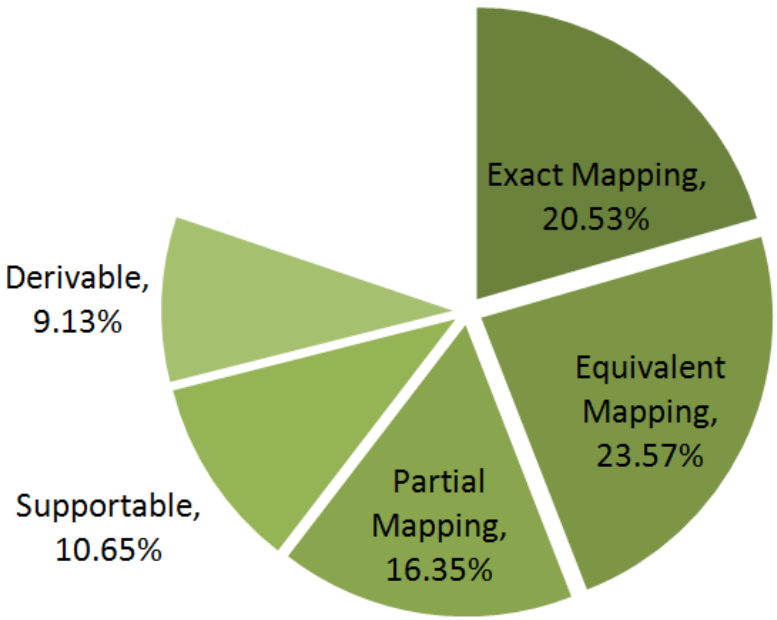

Figure 24. Mapping Details for Harvard Medical School IRB

5.2.4 Western IRB

Western IRB (WIRB) is an independent commercial organization providing IRB services for national and international studies. In our CTSA IRB analysis, I found some institutions required industry-sponsored studies to be reviewed by WIRB. WIRB supports both online applications using “smart” forms and E-mail or hard copy applications using the Word/PDF forms. The Word/PDF forms have the same content design as the Web forms. Seventeen forms were included in the mapping analysis, among them the initial application form, the investigator submission form for multicenter protocols, the initial review submission form for sponsors and contract research organizations (CROs), the initial review submission form for international sites, the screening procedures information form, the Humanitarian Use Device clinical on-label use form, the exemption determination form, forms for request for a partial and full waiver of authorization, the initial review submission form addendum for Department of Defense funded research, the unanticipated problem report form, recruitment bonus disclosure form, the study closure report form, the continuing review report forms, and forms defined for specific states according to local regulations. However, with the most forms and fields (304) defined among the five IRBs, WIRB does not define a standard study protocol form for the IRB application. Investigators and sponsors are required to attach a separate study protocol document along with the application forms. The mapping results for WIRB are shown in Figure 25 and Figure 26.

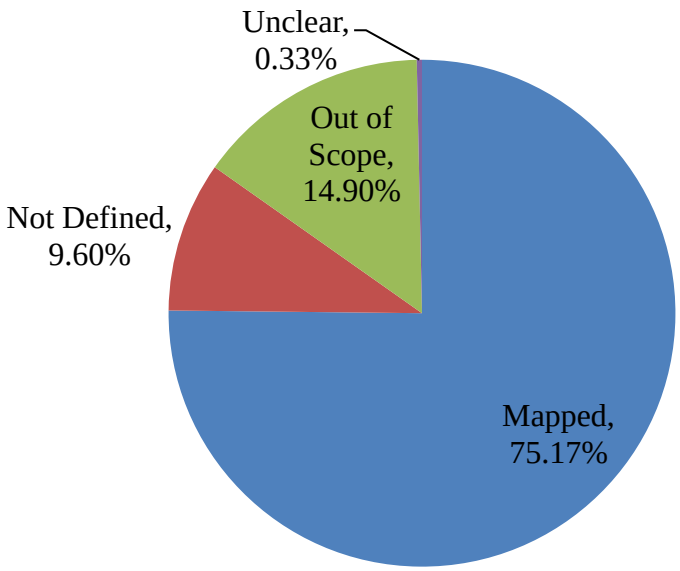

Figure 25. Mapping Results for WIRB

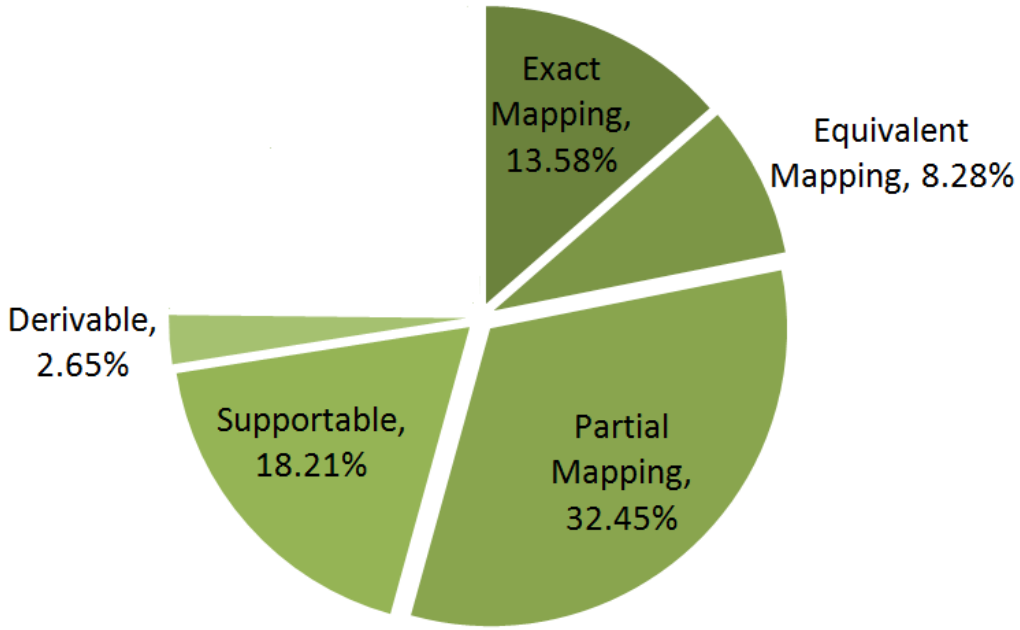

Figure 26. Mapping Details for WIRB

5.2.5 National Cancer Institute's Central IRB (CIRB)

The NCI CIRB Initiative was designed to create a more effective and efficient clinical research effort by conducting full board review of multisite studies centrally, thus reducing the administrative burden on local IRBs and investigators.[90] The CIRB has a clearly defined scope of review, which includes only selected CTEP-sponsored (Cancer Therapy Evaluation Program) Cooperative Group trials. It does not review research involving prisoners, reports of emergency use of a test article, or requests for waivers of HIPAA authorization. The CIRB also reviews individual adverse event reports for studies without a Data and Safety Monitoring Board (DSMB) but it does not define a standard format for the report. Due to the limited scope of studies reviewed by CIRB, the number of fields defined in the CIRB application forms is less than in the application forms from other IRBs (141 fields in total).

The CIRB is switching from its historic facilitated review model to an independent review model. In the independent model, the CIRB is the sole IRB of Record responsible for both study review as well as review of local context considerations for enrolled institutions. Therefore, the CIRB requires information describing local context considerations, which are identified and reported to the CIRB by the Signatory Institutions and Signatory Institutions Principal Investigators via annual and study-specific worksheets.[91]

Both study-specific application forms and annual report sheets defined by CIRB were included in the mapping analysis. A temporary access account was granted by the CIRB to access the online application documents. The mapping results are shown in Figure 27 and Figure 28.

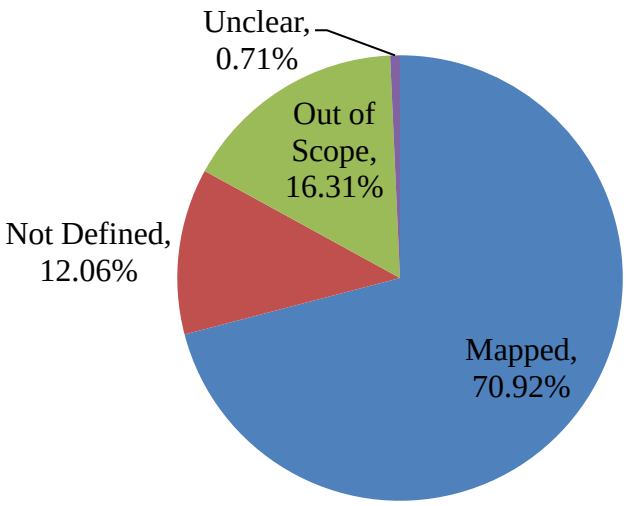

Figure 27. Mapping Results for NCI CIRB

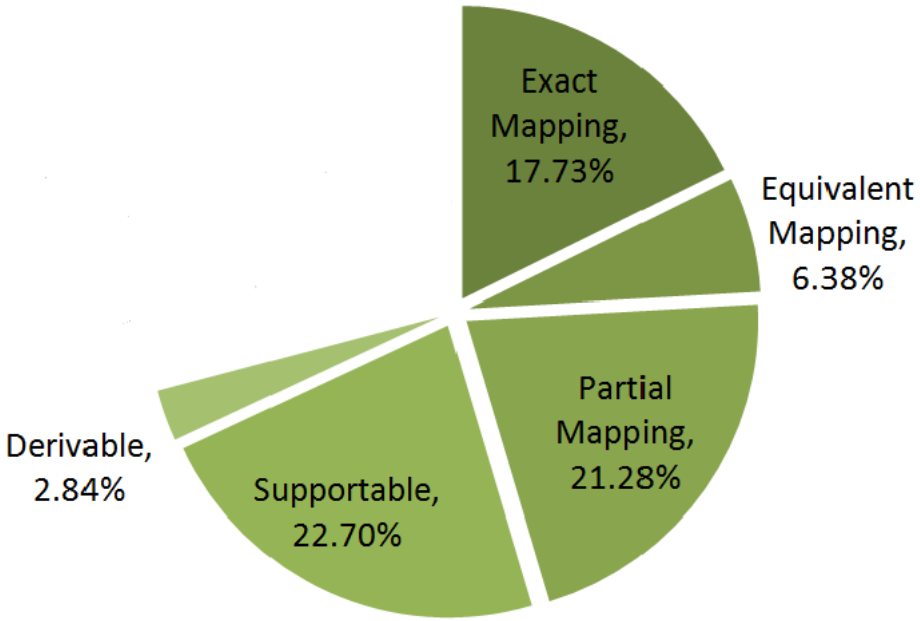

Figure 28. Mapping Details for NCI CIRB

5.3 Discussion

Mapping the fields defined in the above five IRB application systems to the IRB entity-relationship model provides insight into the comprehensiveness of the model and the structure and varieties of real-world IRB application forms. The mapping results can direct future improvements to both model development and IRB application form design. However, since the evaluation was performed by myself, this is a limitation especially considering the complexity of the mapping categories.

The proportion of Exact Mapping is relatively small for all five IRBs but there is great potential to increase it by proper redesign of the Equivalent, Supportable, and Derivable Mapping fields in the application forms. The mapping process and results also show that there is no direct association between electronic IRB and the proportion of Exact Mapping fields. Switching from Word/PDF application forms to e-IRB Web forms does not guarantee more structured application information. For example, WIRB accepts both online applications using its e-IRB system and PDF application forms sent either by mail or electronically. Both submission methods have basically the same content except that the e-IRB system has the "smart" form feature that guides users through the application process depending on users' input. Equivalent Mappings represent the form fields that are currently in free-text format but can be potentially redesigned in a structured format as defined in the Model. Supportable Mappings are valuable sources for future development of the domain vocabulary for the Model since many Supportable Mapping fields suggest the possible values for a certain attribute in the Model. For example, according to the list of options defined in the application renewal forms from some IRBs, the value set for *StatusReport.studyStatus* can be "no

enrollment to date," "enrollment in progress," "no longer enrolling, but active intervention," "no longer enrolling, completed intervention, long-term follow-up," "enrollment permanently closed, only data analysis," "study closed," etc. Derivable Mappings reflect form fields that can be removed to eliminate duplicate manual entries by investigators when they can be automatically populated by the system based on other information items. An example of this is that some IRBs ask specifically about the age of participant, which can be automatically derived from structured eligibility criteria.

In contrast, most Partial Mappings reflect the limitations of using a structured model to cover all information entities in the IRB oversight domain. Partial Mappings represent the form fields that are specified in finer granularity in the real-world IRB application systems but defined using one general attribute in the Model. In most cases, these are the free-text justifications or explanations of different aspects of a single ethical concern (e.g., request for waiver of HIPAA authorization, plan for minimizing the study risks, etc.). These aspects cannot be defined in a structured way in the Model and there is no need to do so since this information is solely for human understanding. Examples of Partial Mappings from WIRB include detailed information items concerning screening procedures (e.g., will a call center be used; how is information stored at the call center; if it is stored in a database, describe the security measures; if it is stored on paper, how and where will the paper be stored and who has access to the paper). All these fields are mapped to *PlannedStudyActivity.activityDescription* when the value for *PlannedStudyActivity*.*reason* is "screen" in the Model.

The Out-of-scope Mapping fields usually represent specific local context considerations and it is not necessary to include them in the core model. However,

individual institutions can decide to include them as local extensions. For example, some IRBs require certain confirmation fields to be checked by investigators that are designed for educational purposes. Specifically, some IRBs ask the investigator who intends to enroll pregnant women to confirm that *1) No inducements will be offered to terminate a pregnancy. 2) Research teams will have no part in decisions related to the timing, method, or procedures used to terminate the pregnancy 3) Research teams will have no part in determining the viability of a neonate*. Such information can be designed as check box fields in the application form, included as instructional or educational information in free-text format, or provided in separate training materials. Among the five IRB application systems, WIRB and CIRB have the most Out-of-scope Mappings because of their different review models. The application forms from these two institutions are designed to consider various local contexts and regulations related to human subject research, which are usually not required explicitly by local IRBs. For example, WIRB requires applicants to attach the clinical trial budget for Canadian sites according to the Canadian Tri-Council Policy Statement. Similarly, WIRB defines information items for New Jersey studies involving inclusion of decisionally impaired subjects because New Jersey Statute 26:14-5 requires investigators to provide additional protections for subjects who are unable to consent for themselves. In addition, WIRB provides informed consent form development and translation services that are not typical among regular IRBs. Form fields like these were considered as out-of-scope for the Model.

However, in other cases, some common local considerations may be appropriate to be included in the Model in the future. For example, many IRBs require investigators to

submit an application even though a study is considered nonhuman subject research by the investigator. Currently the Model does not address the details of this type of study but should be extended as future work.

The Not-Defined form fields reflect elements that are not covered by the current knowledge sources used for developing the Model but are worth considering due to the important role they play in human subject protection. During the mapping analysis, several areas such as humanitarian use device and use of radioactive drugs were identified that are regulated in federal laws but that are not included in AAHRPP guidelines. Section 801 of the Food and Drug Administration Amendments Act (also known as FDAAA 801) requires "applicable clinical trials" to be registered with ClinicalTrials.gov. Although it is not the IRB's responsibility to monitor proper registration of regulated studies, Federal Regulation 21 CFR 50.25(c) has special requirements with regard to the informed consent document for this type of research that should be reviewed carefully by IRB. Therefore, corresponding classes and attributes in these areas were added to the Model after the mapping analysis. There are also undefined form fields in the Model that are based on best practices. For example, some IRBs ask for extra details for placebo-controlled studies. Federal regulations do not address this specific type of study but some IRBs require more information due to potential risks posed by this type of study. A collection of such specializations is valuable to form best practices in the IRB domain and should be discussed in future iterative development of the Model.

The Unclear forms fields should move IRBs to more clearly define fields in their application forms. Ambiguous questions such as "Emergency Drug Information"

defined in an investigational drug data form may cause confusion to researchers. Overly general questions such as “how the risk of the study will be minimized?” or “how the rights and welfare of participants will be protected?” should be avoided.

CHAPTER 6

PROTOTYPE IMPLEMENTATION

This chapter describes a prototype implementation for integrating a SUHD system with an e-IRB system to achieve automated access control on PHI data use for research purposes as proposed in Chapter 2.

6.1 System Overview

The SUHD system chosen is the FURTHeR project as described in Chapter 2. It provides a Web-based data query and export interface (referred to as FURTHeR data access interface or FDAI hereafter) for investigators to directly access aggregated or individual-level health information. The FDAI is adapted from the i2b2 Web front end. The e-IRB system chosen is ERICA, which is the electronic IRB application and review system used at University of Utah. ERICA is commercial software with customized application forms designed by the University of Utah IRB.

The FDAI and ERICA are built on different software platforms, which is not uncommon in an enterprise. The FDAI is an HTML and JavaScript Web application with PhP support, connecting to the Java-based federated query engine at the backend. ERICA is a Web-based portal built on the .NET framework. It also has an add-on module called Click Commerce Extranet that provides a set of Web services enabling external applications to retrieve and manage project information stored in the backend repository.

To integrate these two systems, the Web services technology is a perfect solution since Web services provide a standard means of interoperation between different software applications running on a variety of platforms.[92]

6.2 Interface Specification Design

The first step was to develop an interface specification, which defined all functional requirements to realize the use case scenario illustrated in Figure 20. The IRB entity-relationship model was used as a reference model when developing the input and output parameters of each service interface. The interface specification was reviewed by the ERICA development team. A few rounds of revisions were made based on feedback from this team to accommodate the system's local configurations and limitations. The complete interface specification is listed in Table 7.

The "original query ID" referred to in the service interfaces is the identifier automatically generated by the FURTHeR data access interface for each user-constructed query. In this prototype implementation, the original FURTHeR query ID is sent to ERICA in addition to the actual query (eligibility criteria) message in XML format. This implementation strategy makes later data export service much easier since it only needs to refer to a simple ID to retrieve the query instead of parsing the whole XML query message. The original query ID and the XML query message were designed as hidden fields in ERICA so that they are not visible to end users, which makes the user interface cleaner. Human-friendly inclusion and exclusion criteria are automatically translated from the XML query message and sent to ERICA, but they were designed as non modifiable fields. If users want to modify the eligibility criteria for an IRB application -

Table 7. Interface Specification Developed for the e-IRB-SUHD Integration Based on the IRB DAM

Service Interface Name	**Description**	**Request Parameter**	**Response Result**
getIRBApplicationListByUserId	Retrieve a list of all IRB applications submitted or participated in by a specific user	User ID	A list of IRB applications, each of which contains study summarization information (e.g., study title, PI, original query ID, data access request, application status, etc.)
createIRBApplication	Initiate a new IRB application from the FURTHeR health data access interface	PI ID, original query ID, structured eligibility criteria, free-text inclusion criteria and exclusion criteria, data sources, number of records, data element request	The newly created IRB application ID

Table 7 Continued

Service Interface Name	Description	Request Parameter	Response Result
updateIRBApplication	Update an existing IRB application from the FURTHeR health data access interface	IRB application ID, original query ID, structured eligibility criteria, free-text inclusion criteria, exclusion criteria, data sources, number of records, data element request	Update confirmation message

that was created from the FDAI, they need to update the application using the FDAI. This ensures that the free-text eligibility criteria are always consistent with the original query ID and the corresponding XML query. It can also eliminate the need to build a sophisticated user interface for structured eligibility criteria definition in ERICA, which is not currently supported. A similar strategy was employed for the data element request field, which specifies the data categories and specific data elements to export from FURTHeR. The structured data element request is represented using JSON (JavaScript Object Notation), which is often used as a simplified alternative to XML.

6.3 Implementation and Workflow Demonstration

After achieving consensus on a clearly defined interface specification, the ERICA development team configured the Web services layer to support the functions defined in the interface specification. On the FDAI side, the e-IRB integration was implemented as an i2b2 Web client plug-in. The e-IRB integration plug-in invokes the Web services from ERICA, collects required user input, and implements certain local functions such as query and data request translation, user action auditing, and access control for data export. The following screenshots illustrate the work flow of the integrated system.

1. The investigator logs into the FURTHeR data access interface, composes a data query, specifies the data sources, and runs the aggregated count query (Figure 29).
2. If the investigator is satisfied with the returned count for the study cohort, he or she then may require access to row-level identifiable data for further analysis, which requires IRB approval. In this scenario, the investigator does not have an existing IRB

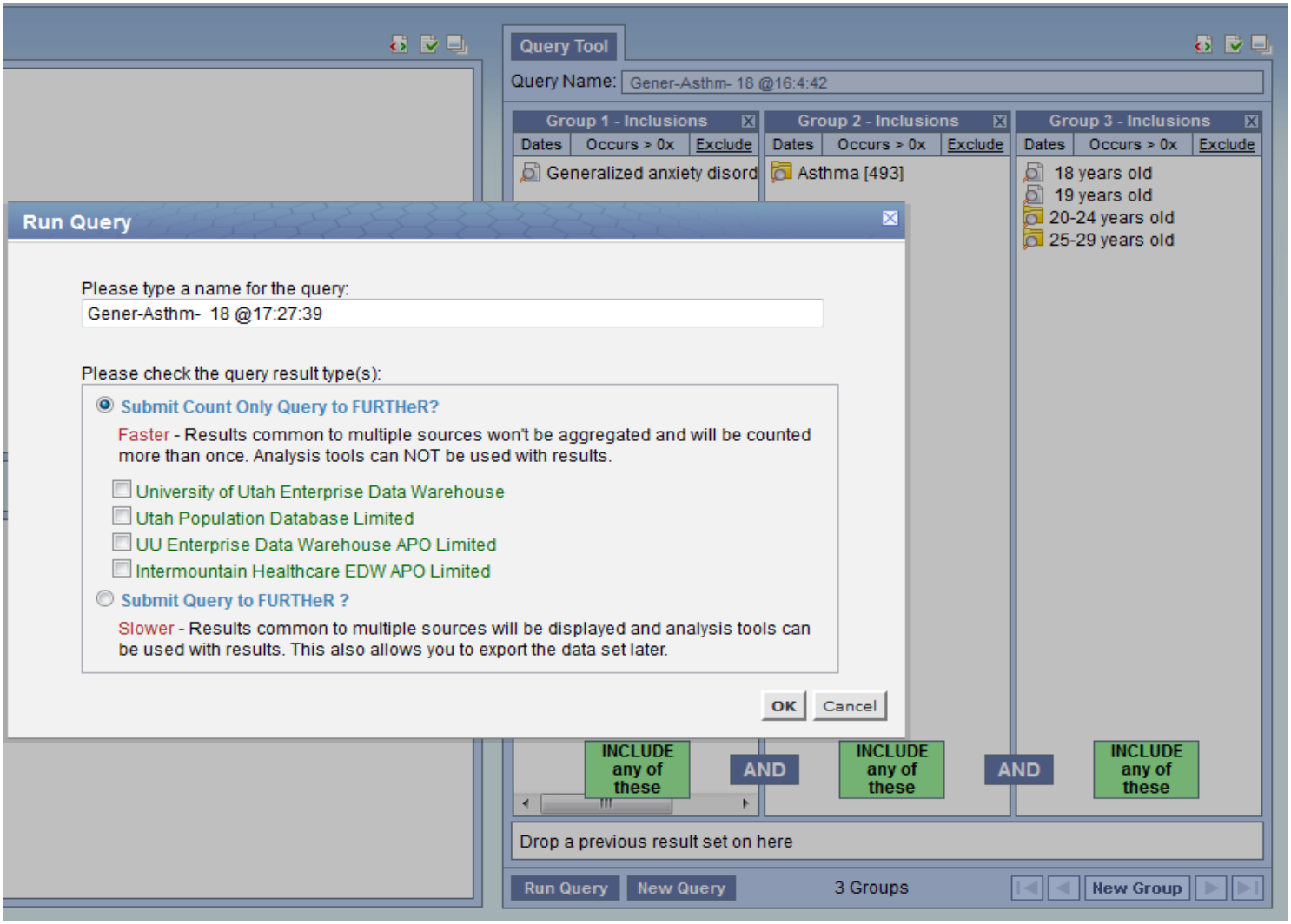

Figure 29. Aggregated Count Query Interface

approval for the row-level data. (The query for aggregate data is considered nonhuman subjects research and does not require IRB approval.)

3. The investigator initiates an IRB application directly from the FDAI using the previous query as inclusion and exclusion criteria and then clicks the "Initiate IRB Application" tab of the e-IRB integration plug-in. The investigator is required to enter some basic study information, drags and drops the previous query, and specifies the data categories (e.g., Demographics, Procedures, Conditions, etc.) and the data elements under each category (e.g., date of birth, race for demographics) he or she wants to access. The data categories and data element names are from the ROHDR model (Figure 30).
4. After filling in the form, the investigator clicks the button "Initiate an IRB application" and all the study related information is sent to ERICA via a Web service, and the newly created IRB application ID is returned. The query composed at the data access interface is represented in XML format (Figure 31), which is unlikely to be comprehensible to IRB reviewers. A program was developed to translate the XML query into a human understandable format represented by inclusion and exclusion criteria (Figure 32). A similar strategy applies to the data element request that uses JSON as the machine interpretable format.
5. The investigator clicks the ERICA link and is directed to the ERICA Web portal to finish the rest of the application, with data request related fields collected from FURTHeR automatically populated. (Note: Single Sign-On is not supported in ERICA now but it can be easily configured in the future so the user will be directly taken to the ERICA application without the extra click.)

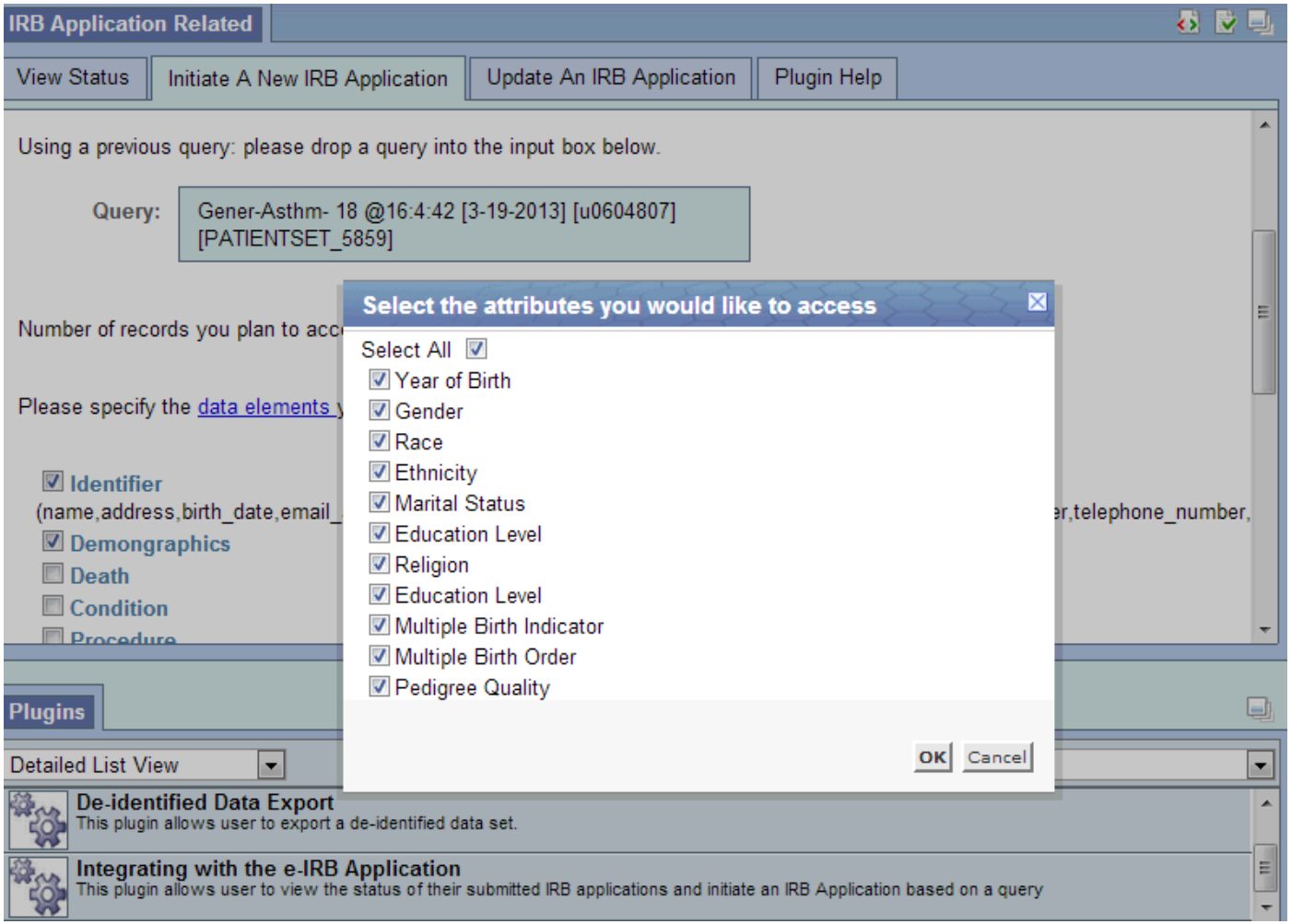

Figure 30. Interface for Specifying Data Access Request

```
<panel>
    <panel_number>1</panel_number>
    <invert>0</invert>
    <total_item_occurrences>1</total_item_occurrences>
    <item>
        <hlevel>7</hlevel>
        <item_name>Generalized anxiety disorder [300.02]</item_name>
        <item_key>\\FURTHER\FURTHER\ICD-9\Diseases and injuries (001-999.99)\Mental disorders
        (290-319.99)\Neurotic disorders, personality disorders, and other nonpsychotic mental disorders
        (300-316.99)\Anxiety, dissociative and somatoform disorders (300)\Anxiety states (300.0)\Generalized
        anxiety disorder (300.02)\</item_key>
        <tooltip>ICD-9 \ Diseases and injuries (001-999.99) \ Mental disorders (290-319.99) \ Neurotic
        disorders, personality disorders, and other nonpsychotic mental disorders (300-316.99) \ Anxiety,
        dissociative and somatoform disorders (300) \ Anxiety states (300.0) \ Generalized anxiety disorder
        (300.02)</tooltip>
        <class>ENC</class>
        <constrain_by_date>
        </constrain_by_date>
        <item_icon>LA</item_icon>
        <item_is_synonym>false</item_is_synonym>
    </item>
</panel>
<panel>
    <panel_number>2</panel_number>
```

Figure 31. The XML Representation of the Eligibility Criteria Generated by the FDAI

5. Characteristics of Participants/Inclusion Criteria:

HELP?

Participant-entry criteria should be as detailed as necessary to define the participant population under study and, for clinical studies, to reduce confounding treatments or diseases. Precise criteria for age, gender, or another other factors (e.g. diagnoses, extremes in signs or symptoms, etc.) should be included.

Generalized anxiety disorder [300.02];

AND

Asthma [493];

AND

18 years old OR 19 years old OR 20-24 years old OR 25-29 years old;

Figure 32. The Automatically Translated Eligibility Criteria in a Human-friendly Format

6. The investigator can obtain a quick view of the submitted IRB applications via the e-IRB integration plug-in. He or she can also modify the query and data element request that were submitted with an IRB application. If the IRB application is approved, the data export function will be enabled. The investigator may click on the "Export the data" link in an IRB application status page (Figure 33) to run the approved query and obtain the data. Since the IRB application is stored with the original query ID and structured data element request, FURTHeR query engine can easily interpret such information to return the data set approved by IRB, thus realizing automated access control based on IRB approval.
7. A log table is created in the FURTHeR end to store user-initiated actions related to data export for future analysis or audit (Figure 34).

6.4. Summary of the Prototype Implementation

Currently the FURTHeR query engine does not yet support PHI query and federation. This prototype implementation demonstrates the feasibility from the workflow perspective in the front end. The development work was performed on test instances of the data access interface and the ERICA system to avoid interrupting the production users.

The front-end implementation using an i2b2 plug-in can be easily adapted by other institutions that already deployed the i2b2 suite as their data management solution. According to the i2b2 Website, there are over 50 institution users of the i2b2, including over half of the CTSAs.[93] This means that our IRB-based access control can be generalized to these i2b2 adopters as long as they have an electronic IRB.

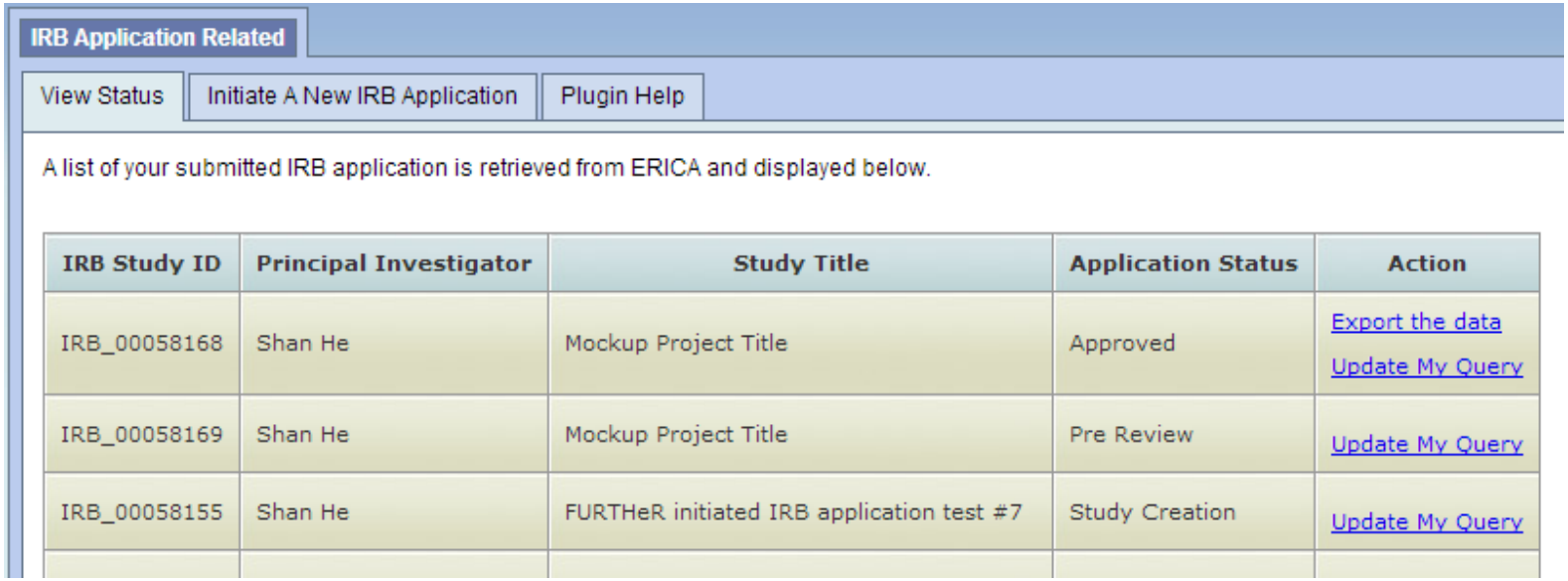

Figure 33. Example User Interface for Viewing Submitted IRB Application via FDAI

log_id	action_code	irb_study_id	i2b2_query_id	data_request	action_date
1	Initiate IRB Applica	IRB_00058172	5601	{"demographics":&	2013-03-25
2	Data Export	IRB_00058172	5601	","pathologyObserv	2013-03-28

Figure 34. Log Table for the e-IRB Plug-in

CHAPTER 7

DISCUSSION AND CONCLUSION

7.1 Summary

This dissertation analyzed the problems associated with information technology use in the IRB oversight (ethics review) domain. Over 100 IRB application systems used at CTSA institutions were analyzed with regard to their submission method (e.g., paper, e-mail, or online) and the sources of online systems (e.g., vendor, in-house built, SaaS, community source, etc). The lack of standardized IRB application forms caused inefficient and inconsistent review and cumbersome workflow due to disconnect between systems. The need for building a domain analysis model for IRB oversight was also discussed. The goal of developing an IRB domain analysis model is to standardize the structure and content of IRB application forms and promote system interoperability among CRI systems to streamline the clinical research process.

This dissertation contains a comprehensive literature review on existing domain analysis models and ontologies that are related to the IRB oversight domain. Although some of the existing models or ontologies overlap with the IRB domain to a certain extent, no previous research specifically addressed the modeling issue in this domain. A domain analysis process was designed in this dissertation to address the major goal of building an IRB DAM based on several domain analysis methods from software engineering. Previous modeling research reported that domain experts have problems

understanding UML diagrams, which was also encountered in this project. Concept map technology was used for domain expert understanding and review of the model, which is a lesson learned during the modeling process and can be applied to other modeling research projects. The final IRB DAM includes the structural entity-relationship model and the behavioral business process model, as well as the interaction architecture model. The IRB entity-relationship model (information model) is the most important component of the entire IRB DAM and was evaluated by a comparison with five real-world IRB application systems. The evaluation results were discussed in detail. A prototype implementation of the model integrated the FURTHeR data access interface with the e-IRB system ERICA in a testing environment. It demonstrated automated access control on PHI based on IRB approval.

Development of the IRB DAM was initially motivated by addressing the need to integrate a health data query system and an e-IRB system to realize automated access control on PHI based on IRB approval. However, the value of the IRB DAM extends beyond this use case. Structured and computable IRB application information can facilitate automated review decision support with predefined rules, thus enhancing the review quality and efficiency. More importantly, the IRB DAM can be used as the basis for data exchange message development between e-IRB systems and various CRI systems. Effective exchange and sharing of study data and metadata among various CRI systems is the key to streamline the clinical research process. A weak link in the chain at any point can cause unnecessary waste of resources and time. Development of the IRB DAM fills a gap in standardization and modeling efforts in the IRB oversight domain in the clinical research community.

7.2 Limitations

The model validation process revealed limitations of the IRB DAM in representing certain aspects of real-world IRB applications, especially in defining information elements about subjective evaluations and justifications from investigators for a certain study activity, or foreseen events that can be fully expressed only with free text. This limitation is caused by the nature of information models whose strength lies in representing discrete and machine-understandable data elements, but not free-text information. Electronic IRB systems designed with structured application forms and predefined review rules cannot replace human review. The goal of building the IRB DAM is to promote standardization in IRB application forms that can enhance more efficient and consistent ethics review across IRBs and improve system integration to streamline the clinical research workflow.

I performed the evaluation of the model myself, which may cause some mapping biases. However, the actual value of the evaluation does not lie in the specific numbers listed in the mapping result table (Table 6). As demonstrated by the complicated mapping results, the purpose of the evaluation is not to categorize the model simply as "good" or "bad." There is no gold standard regarding to the design of IRB application forms. The 5 IRB application systems chosen in the evaluation phase are representative across the nation, but this does not mean their application forms are perfect. The comparison between the IRB model and the real-world IRB application forms identified 1) if the model has covered the core information elements required by IRBs; 2) for information that are not covered in the model, what belong to local context and should not be included in the core model and what may be potentially considered in the future version

of the model according to best practice; 3) what information that are currently free texts can be defined in a structured format according to the model; 4) what fields in the current IRB application forms should be clarified. In short, the evaluation provided insight with regard to future improvement to both the IRB model development and real-world IRB application form design.

This dissertation only did a prototype implementation of the model in a testing environment. Although the demonstrated integration of ERICA and FURTHeR showed a streamlined workflow for PHI data access, a formal evaluation of performance improvement and user satisfaction of the integrated workflow is needed to make the proposed solution more convincing. However, implementing the integration between the two systems in the production environment will involve complicated administrative procedure change and political discussions, which is considered as future work.

7.3 Future Direction

Like any modeling effort, development of an IRB DAM that meets real-world application requirements needs many rounds of iteration and revision. As future work, we envisage continuing iterative development of the model by collaborating with more IRB domain experts and clinical researchers. A formal evaluation of the expressiveness of the data request related classes needs to be performed, possibly by annotating previously submitted IRB applications.

We plan to promote the adoption of the IRB model by collaborating with other CTSA centers that have e-IRB implementations according to the analysis result in Chapter 3. Although the e-IRB system in our prototype implementation was integrated with a SUHD system, it can be integrated with any clinical research information system such as CTMS

and EDC systems, which will be part of the future work to demonstrate the implementation of the IRB model. At the same time, we plan to integrate the IRB DAM with the BRIDG model through a harmonization process developed by BRIDG. BRIDG is a well-known modeling effort that has a great impact in the clinical research domain across the nation. Integrating our model with BRIDG will complement the regulatory artifacts in BRIDG and raise the awareness of our model among other institutions.

A comprehensive domain vocabulary specification for the IRB DAM needs to be developed to achieve semantic interoperability. This can be achieved by combining the validation results from Chapter 5 (especially those Supportable Mapping fields), domain expert interviews, and existing terminologies in the healthcare and clinical research domain. In addition, computable review rules can be developed based on the structured information model and applied to e-IRB systems to facilitate ethics review decision support. More advanced features such as computer-assisted consent form generation based on study procedures and risks can also be realized at a later stage of the model adoption.

REFERENCES

1. Kahn MG, Weng C. Clinical research informatics: a conceptual perspective. JAMIA. 2012 Jun;19(1e):e36–42.

2. Embi PJ, Payne PRO. Clinical research informatics: challenges, opportunities and definition for an emerging domain. JAMIA. 16(3):316–27.

3. REDCap [Internet]. [cited 2013 Feb 12];Available from: http://project-redcap.org/

4. OpenClinica [Internet]. [cited 2013 Feb 12];Available from: https://community.openclinica.com/

5. Electronic Submissions Gateway [Internet]. [cited 2013 Feb 12];Available from: http://www.fda.gov/ForIndustry/ElectronicSubmissionsGateway/default.htm

6. MedWatch: The FDA Safety Information and Adverse Event Reporting Program [Internet]. [cited 2013 Feb 12];Available from: http://www.fda.gov/Safety/MedWatch/default.htm

7. Title 45 Public Welfare Department of Health and Human Services Part 46 Protection of Human Subjects [Internet]. [cited 2013 Feb 13];Available from: http://www .hhs.gov/ohrp/humansubjects/guidance/45cfr46.html

8. Title 21 Food and Drugs Part 50 Protection of Human Subjects [Internet]. [cited 2013 Feb 13];Available from: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?cfrpart=50

9. Title 21 Food and Drugs Part 56 Institutional Review Boards [Internet]. [cited 2013 Feb 13];Available from: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?cfrpart=56

10. Title 21 Food and Drugs Part 312 Investigational New Drug Application [Internet]. [cited 2013 Feb 13];Available from: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?cfrpart=312

11. Title 21 Food and Drugs Part 812 Investigational Device Exemptions [Internet]. [cited 2013 Feb 13];Available from: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?cfrpart=812

12. U.S. Department of Health & Human Service Omnibus HIPAA Rulemaking [Internet]. [cited 2013 Feb 13];Available from: http://www.hhs.gov/ocr/privacy/hipaa/administrative/omnibus/

13. Catania JA, Lo B, Wolf LE, Dolcini MM, Pollack LM, Barker JC. Survey of U.S. Human Research Protection Organizations: Workload and Membership. J Empir Res Hum Res Ethics. 2008;I(4):57–69.

14. Trends, Charts and Maps [Internet]. [cited 2013 Feb 13];Available from: http://clinicaltrials.gov/ct2/resources/trends#RegisteredStudiesOverTime

15. Weinberger M, Oddone EZ, Henderson WG, Smith DM, Huey J, Giobbie-Hurder A, et al. Multisite randomized controlled trials in health services research: scientific challenges and operational issues. Medical Care. 2001 Jun;39(6):627–34.

16. McWilliams R, Hoover-Fong J, Hamosh A, Beck S, Beaty T, Cutting G. Problematic variation in local institutional review of a multicenter genetic epidemiology study. JAMA. 2003 Jul 16;290(3):360–6.

17. Burman WJ, Reves RR, Cohn DL, Schooley RT. Breaking the camel's back: multicenter clinical trials and local institutional review boards. Annals of Internal Medicine. 2001 Jan 16;134(2):152–7.

18. Greene SM, Geiger AM. A review finds that multicenter studies face substantial challenges but strategies exist to achieve Institutional Review Board approval. Journal of Cinical Epidemiology. 2006 Aug;59(8):784–90.

19. Graham DG, Spano MS, Manning B. The IRB challenge for practice-based research: strategies of the American Academy of Family Physicians National Research Network (AAFP NRN). JABFM. 20(2):181–7.

20. Ness RB. Influence of the HIPAA Privacy Rule on Health Research. JAMA. 2007 Nov 14;298(18):2164–70.

21. Straight TM. Clinical research regulation: challenges to the institutional review board system. Clinics in Dermatology. 2009;27(4):375–83.

22. Silberman G, Kahn KL. Burdens on research imposed by institutional review boards: the state of the evidence and its implications for regulatory reform. The Milbank Quarterly. 2011 Dec;89(4):599–627.

23. Stiroh KJ. Information Technology and the U . S . Productivity Revival : What Do the Industry Data Say ? Federal Reserve Bank of New York. 2001.

24. Microsoft SharePoint [Internet]. [cited 2013 Mar 1];Available from: http://office.microsoft.com/en-us/sharepoint/sharepoint-online-online-collaboration-software-FX103789366.aspx

25. Brown B, Merritt MW. A global public incentive database for human subjects research. IRB. 35(2):14–7.

26. Gold JL, Dewa CS. Institutional review boards and multisite studies in health services research: is there a better way? Health Services Research. 2005 Feb;40(1):291–307.

27. Kuchinke W, Aerts J, Semler SC, Ohmann C. CDISC standard-based electronic archiving of clinical trials. Methods of Information in Medicine. 2009 Jan;48(5):408–13.

28. Freimuth RR, Freund ET, Schick L, Sharma MK, Stafford GA, Suzek BE, et al. Life sciences domain analysis model. JAMIA. 19(6):1095–102.

29. Livne OE, Schultz ND, Narus SP. Federated querying architecture with clinical & translational health IT application. Journal of Medical Systems. 2011 Oct;35(5):1211–24.

30. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). JAMIA. 2010;17(2):124–30.

31. Narus SP, Schultz ND, Livne OE, Bradshaw RL, Mitchell JA. Going FURTHeR with i2b2. In: AMIA Annual Symposium Proceedings. 2011.

32. Reisinger SJ, Ryan PB, O'Hara DJ, Powell GE, Painter JL, Pattishall EN, et al. Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. JAMIA. 17(6):652–62.

33. Alemayehu D, Mardekian J. Infrastructure requirements for secondary data sources in comparative effectiveness research. JMCP. 2011;17(9 Suppl A):S16–21.

34. Köpcke F, Kraus S, Scholler A, Nau C, Schüttler J, Prokosch H-U, et al. Secondary use of routinely collected patient data in a clinical trial: An evaluation of the effects on patient recruitment and data acquisition. International Journal of Medical Informatics. 2013 Mar;82(3):185–92.

35. Weng C, Batres C, Borda T, Weiskopf NG, Wilcox AB, Bigger JT, et al. A real-time screening alert improves patient recruitment efficiency. In: Annual Symposium Proceedings. 2011. p. 1489–98.

36. Langella S, Hastings S, Oster S, Pan T, Sharma A, Permar J, et al. Sharing data and analytical resources securely in a biomedical research Grid environment. JAMIA. 2008;15(3):363–73.

37. Murphy SN, Mendis M, Hackett K, Kuttan R, Pan W, Phillips LC, et al. Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside. In: AMIA Annual Symposium Proceedings. 2007. p. 548–52.

38. S&I Framework Health Query [Internet]. [cited 2013 Mar 1];Available from: http://wiki.siframework.org/Query+Health

39. Rea S, Pathak J, Savova G, Oniki T a, Westberg L, Beebe CE, et al. Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPn project. JBI. 2012 Aug;45(4):763–71.

40. Weber GM, Murphy SN, McMurry AJ, Macfadden D, Nigrin DJ, Churchill S, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. JAMIA. 2009;16(5):624–30.

41. Oregon Clinical and Translational Research Institute Cohort Discovery [Internet]. [cited 2013 Jan 16];Available from: http://www.ohsu.edu/xd/research/centers-institutes/octri/resources/octri-research-services/cohort-discovery.cfm

42. Stanford Center for Clinical Informatics STRIDE Cohort Discovery Tool [Internet]. [cited 2013 Jan 16];Available from: https://clinicalinformatics.stanford.edu/services/cohort.html

43. Klann JG, Murphy SN. Query Health and i2b2 : Enabling Standards-based , Multiplatform Population Health Queries. 2012. p. 90.

44. He S, Hurdle JF, Botkin JR, Narus SP. Integrating a Federated Healthcare Data Query Platform With Electronic IRB Information Systems. AMIA Annual Symposium Proceedings. 2010 Jan;2010:291–5.

45. U.S. Department of Health and Human Services. Advanced Notice of Proposed Rulemaking for Revision to the Common Rule. 2011.

46. Sim I, Carini S, Tu SW, Detwiler LT, Brinkley J, Mollah SA, et al. Ontology-based federated data access to human studies information. In: Annual Symposium Proceedings. 2012. p. 856–65.

47. CDISC. The Protocol Representation Model Version 1.0 [Internet]. 2010 [cited 2013 Jan 16];Available from: http://www.cdisc.org/protocol

48. Fridsma DB, Evans J, Hastak S, Mead CN. The BRIDG project: a technical report. JAMIA. 2008;15(2):130–7.

49. The OBI Consortium [Internet]. [cited 2013 Mar 3];Available from: http://obi-ontology.org/

50. Koepsell D, Arp R, Smith B. Creating a Controlled Vocabulary for the Ethics of Human Research: Towards a Biomedical Ethics Ontology. 2010;4(1):43–58.

51. Dubois JM. The Biomedical Ethics Ontology Proposal: Excellent Aims, Questionable Methods. 2010;4(1):59–62.

52. IRB Guidebook - U.S. Department of Health and Human Services [Internet]. [cited 2013 May 4];Available from: http://www.hhs.gov/ohrp/archive/irb/irb_chapter6.htm

53. Mackenzie SL, Wyatt MC, Schuff R, Tenenbaum JD, Anderson N. Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey. JAMIA. 2012 Jun 1;19(e1):e119–e124.

54. OnCore eClinical Solution: Improving Clinical Trial and Research Management [Internet]. [cited 2013 Apr 3];Available from: http://forteresearch.com/eclinical-oncore/

55. Harris P a, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. JBI. 2009 Apr;42(2):377–81.

56. REDCap [Internet]. [cited 2013 Apr 3];Available from: http://project-redcap.org/

57. Kuali Coeus [Internet]. [cited 2013 Feb 5];Available from: http://www.kuali.org/kc

58. Neighbors JM, Freeman PA, Gostelow KP, Tonge FM. Software Construction Using Components. 1980;:1–75.

59. Prieto-diaz. Domain Analysis: An Introduction. 1990.

60. Prieto-Diaz R. Domain Analysis for Reusability. 1987.

61. Ferré X. An Evaluation of Domain Analysis Methods.

62. Embley DW, Thalheim B. Handbook of Conceptual Modeling: Theory, Practice, and Research Challenges. 2011.

63. HL7 Domain Analysis Model [Internet]. [cited 2013 Feb 2];Available from: http://wiki.hl7.org/index.php?title=Domain_Analysis_Model

64. Cohen SG, Hess JA, Novak WE, Peterson AS. Feasibility Study Feature-Oriented Domain Analysis (FODA) Kyo C . Kang. 1990;(November).

65. Guidance on Reviewing and Reporting Unanticipated Problems Involving Risks to Subjects or Others and Adverse Events.

66. Rumbaugh J, Jacobson I, Booch G. The Unified Modeling Language Reference Manual. 1999.

67. Giantdino B, Booch G, Rumbaugh J, Jacobson I, Matter F. The Unified Modeling Language User Guide. 1998.

68. The Association for the Accreditation of Human Research Protection Programs, Inc. (AAHRPP) [Internet]. [cited 2013 Mar 1];Available from: http://www.aahrpp.org/

69. Evaluation Instrument for Accreditation [Internet]. 2012 [cited 2012 Aug 1];2012. Available from: http://www.aahrpp.org/apply/resources/evaluation-instrument-for-accreditation

70. IRB Evaluation Checklist [Internet]. 2010 [cited 2012 Aug 1];Available from: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&ved=0CC0QFjAA&url=https%3A%2F%2Fadmin.share.aahrpp.org%2FWebsite%2520Documents%2FIRB_Evaluation_Checklist.DOC&ei=fjHkUbCUHKmFiAKp-4HICw&usg=AFQjCNEMaFAVuyRw_DXdi90z0WRIKKyhpw&sig2=qZsDbXqlHkytgabQ2Pmb4g&bvm=bv.48705608,d.aWM

71. U.S. DHHS Office for Human Research Protections (OHRP) [Internet]. [cited 2013 Feb 1];Available from: http://www.hhs.gov/ohrp/

72. Clinical Research and the HIPAA Privacy Rule. 2004.

73. Institutional Review Boards and the HIPAA Privacy Rule. 2003.

74. Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule. 2004.

75. Research Repositories, Databases, and the HIPAA Privacy Rule. 2004.

76. Novak JD, Cañas AJ. The Theory Underlying Concept Maps and How to Construct and Use Them 1. 2008;:1–36.

77. OMG. Business Process Model and Notation (BPMN) Version 2.0. 2011.

78. 21 CFR 56.102 Definitions (d) [Internet]. [cited 2013 Mar 1];Available from: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=56.102

79. FDA Humanitarian Device Exemption [Internet]. [cited 2013 Feb 1];Available from: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/HumanitarianDeviceExemption/

80. The Human Studies Database Wiki [Internet]. 2012 [cited 2013 Feb 1];Available from: http://hsdbwiki.org/index.php/HSDB_Collaborative_Wiki

81. HL7 Version 3 Standard: Data Types - Abstract Specification, Release 2 [Internet]. [cited 2013 May 6];Available from: http://www.hl7.org/implement/standards/ product_ brief .cfm?product_id=264

82. Tu SW, Peleg M, Carini S, Bobak M, Ross J, Rubin D, et al. A practical method for transforming free-text eligibility criteria into computable criteria. JBI. 2011 Apr;44(2):239–50.

83. Weng C, Tu SW, Sim I, Richesson R. Formal representation of eligibility criteria: a literature review. JBI. 2010 Jun;43(3):451–67.

84. Patel CO, Weng C. ECRL: an eligibility criteria representation language based on the UMLS Semantic Network. AMIA Annual Symposium Proceedings. 2008 Jan;:1084.

85. Weng C, Wu X, Luo Z, Boland MR, Theodoratos D, Johnson SB. EliXR: an approach to eligibility criteria extraction and representation. JAMIA. 2011 Dec;18 Suppl 1:i116–24.

86. IEEE. Standard Computer Dictionary [Internet]. 1991 [cited 2013 Mar 1];Available from: http://ieeexplore.ieee.org/servlet/opac?punumber=2267

87. Bakken S, Campbell KE, Cimino JJ, Huff SM, Hammond WE. Toward vocabulary domain specifications for health level 7-coded data elements. JAMIA. 7(4):333–42.

88. Goossen W, Goossen-Baremans A, van der Zel M. Detailed clinical models: a review. Healthcare Informatics Research. 2010 Dec;16(4):201–14.

89. Observational Medical Outcomes Partnership Common Data Model Specifications. 2012;(April).

90. NCI CIRB Standard Operating Procedures. 2013.

91. Synopsis of the CIRB Independent Model. 2012.

92. Web Services Architecture. 2004.

93. Collaborations: i2b2 Installations [Internet]. [cited 2013 Mar 3];Available from: https://www.i2b2.org/work/i2b2_installations.html